# Model Calculation Tool – Guide

## Overview

The Model Calculation Tool (MCT) is a structure by which the concepts proposed and discussed in the Pay-It-Forward (PIF) research project can be put into practice on a real data set. At a high level, the MCT takes as its input the publication data and library budget data from a particular institution, allows the user to manipulate various model parameters discussed throughout the PIF report, and returns the anticipated cost allocation to the various stakeholders (libraries, via subsidies; granting agencies when available; and other discretionary funding to which the author has access).

With the MCT, users can observe how the proposed financial model would impact a specific institution, and can observe how various changes in environment (such as percentage of research with grant funding available or growth in publication) as well as various institution-specific options (such as the choice of a library subsidy or the decision of whether to offer a subsidy to all authors or just those without grants) affect the outcome of the model.

## Data input

Raw data used for the MCT is gathered and refined according to the methods described in the PIF report. For the institutions and years within the scope of the PIF study, these data inputs can be readily generated from our raw data files. For other institutions, data can be gathered and input according to similar methods.

### Publication data

A dataset representing publication output for a single year at a single institution is generated from gathered data. Data can be included at the article level, aggregated by some factor (such as journal), or a combination of the two. This raw data set should first be generated outside of the MCT, and then pasted in to columns A through I of the “Raw Data” tab of the MCT. Finally, clicking the “Complete Data Import” applies the appropriate analysis fields to each line of data, finalizing the data import.

This dataset includes several required components and several optional components:

1. **PIF\_Subject** (*required*): the PIF subject of this paper or set of papers, used in applying subject-specific parameters and in observing subject-specific result sets. Currently, use of the PIF subject scheme is required for this tool.
2. **DOCTYPE** (*optional*): the standardized document type for this paper or set of papers. Possible values are Article, Review, or Proceedings. Currently listed as optional because there are no parameters referring to document type yet, but there may be in the future.
3. **DataLevel** (*optional*): the aggregation level of this line of data; possible values include Article and Journal. The data in columns H and I mean that this column is not necessary, but may be useful for troubleshooting data issues if necessary.
4. **DataSource** (*required*): the source of this particular line of data. This column is required because papers without a SNIP value available for calculation of the APC instead use the average APC of all papers from that particular data source.
5. **SourceID** (*optional*): an identifier referring to the journal or other aggregation level for this line of data (e.g. ISSN, Scopus source\_id). Optional, but may be useful for troubleshooting data issues if necessary. Can be set as “NA” when the ArticleID column provides an identifier instead.
6. **ArticleID** (*optional*): an identifier referring to this particular paper (e.g. DOI, WoS accession number). Optional, but may be useful for troubleshooting data issues if necessary. Can be set as “NA” when the SourceID column provides an identifier instead.
7. **JournalValueMetric** (*required*): Any numerical journal value metric; the metric used in the PIF report and all examples and sample data is the 2014 SNIP value, as assigned at <https://www.journalmetrics.com/>. In the PIF report we build our model under the assumption that journal publishers set APCs relative to the value that authors reap from publishing in their journals. Therefore, this data point is used to predict the APC that authors will pay for this paper (or for a paper in this aggregated set), based on the linear equation defined in the APC Parameters tab (see below).
8. **ExpectedDocs** (*required*): The number of papers from this data line for which we expect the journal to charge the APC to an author at this institution; in our modeling this means the number of papers where the corresponding author is at the institution.

For actual article-level data where it is known that the journal will charge the APC to this institution, this data point is 1. For aggregated data, this data point represents the number of articles where the given journal will charge APCs to this institution. This data point can also be a decimal, if the data is the result of applying probabilities or expected values. For example, say a journal published 4 articles with institutional authors, but data is not available regarding which of those papers the journal will charge the APC to this institution. The MCT user could estimate what percentage of APCs will be charged to the institution (say, 60%), and apply that estimate to the total number of papers; in this example, the ExpectedDocs data point would be 2.4. Note that this calculation should be done outside of the MCT: data in this field should only represent the number of APCs where this institution will be responsible for payment.
9. **ExpectedGrantDocs** (*required*): the number of papers from this data line which have grant funding available to them (or, which acknowledged a grant in the paper). This value should be less than or equal to column H, ExpectedDocs.

For actual article-level data, this data point would likely be 1 or 0, representing whether the paper has a grant or not. As with the ExpectedDocs data point, for aggregated data, this data point can be higher than 1, and it could be a decimal if the data is the result of applying probabilities or expected values. For example, if grant acknowledgement data is unavailable, but we expect that 75% of papers in the given subject are the result of grant-funded research, this data point would be 75% of the value in the ExpectedDocs column.

Anticipated grant percentages can also be modeled in the Advanced Parameters tab; if used, those values will override the data entered here.

### Model data

This sheet contains a small set of data about the model itself, used mainly for sustainability comparisons and presentation. These data points include:

* **Redirectable library expenditures** represents the total amount of money in the given year the library spent on resources which will be freely available under an APC-funded model. This should include subscription expenditures as well as library-funded APC funds and memberships with OA publishers.
* **Total extramural research expenditures** represents the total amount of money from external sources (i.e. grant agencies) spent on research in the given year. In our data set, for U.S. institutions, this is gathered from the HERD survey, with Institutional Sources removed from the total.
* **Institution** is the university being examined here
* **Year** is the year under discussion; for publications this means calendar year (January X to December X), and for finances this means fiscal year (July X-1 to June X)

## Parameter selection

The parameter selection sheets allow for a choice of model inputs or parameters based on expected environmental or economic conditions, as well as based on choices that the library or institution might make. These include components such as how publishers assign an APC, how much of a subsidy libraries offer to authors, what expectations institutions put on their authors for using grant funds, what percentage of the literature actually adopts the APC-funded model, and how various parameters grow over time. The analysis of the publication data and model data is then performed based on the given parameters. In general, anything colored in light green is a parameter that can be changed to alter the calculated outcome of the model.

### Overview/Cost equation

The analyses performed in the MCT are based on two main concepts discussed within the PIF report:

#### Total Cost Equation

The total cost equation, proposed in the “Financial Model” section of the PIF report (p.91), governs the expected cost to the institution for a year’s worth of publishing research articles. The equation is defined as:

**APCtotal = PUB x PA x PR x APCavg x (1 + AG)y x (1 + APCI)y**

Each variable in the equation is represented by either the publication data input into the system above (**PUB** and **PR**), or one of the parameters described below (all others). Based on this data and these parameters, the total cost to the institution for this year of publication is determined. Furthermore, as is discussed in the report, the **APCavg** factor can be split into components representing the portion of each APC assigned to each stakeholder. How these components are determined is also defined by the parameters below, and based on that determination, the total cost allocated to each stakeholder can be calculated. All calculated costs can also be broken out by discipline, and can be compared with various other measures to assess potential sustainability.

#### APC pricing under a competition model

As discussed in the “Estimating APC Pricing” section, under a model where competition for authors is introduced, publishers are expected to base the APC they charge for publishing in a particular journal on the value that authors gain from publishing in that journal. In our analysis, we use Source Normalized Impact per Paper as a proxy for this value, and we use a linear regression to develop an equation estimating the APC for a specific journal based on its SNIP value:

***APC* = 1147.68 + 709.4 x SNIP**

The specifics of how we developed this equation are discussed in the report (p.102-103). However, we acknowledge that an improved data set or slightly altered assumptions about how to calculate this equation could result changes to this equation, so the various parameters of the APC assignment are included below as well. This equation is applied to every paper in the publication data set to calculate the APC that the institution’s authors will pay.

### APC parameters

This worksheet allows users to redefine the equation used to estimate the APC that a journal will charge based on the journal value metric (JVM) assigned in the raw data, in column G. The default equation, defined above, is based on an analysis of OA journal list-price APCs and SNIP values. This equation can be generalized to any choice of JVM:

$$APC= \left\{\begin{array}{c}X+Y×JVM for 0\leq JVM\leq Z \\MaxAPC for JVM>Z\end{array}\right.$$

There are four parameters to set on this tab, listed below. It is recommended that the first three parameters are only changed as the result of a new, updated, or alternative analysis; the fourth could be modified based on a different understanding or observation of common practice.

* **Intercept** is the minimum APC assigned to any article, applicable when the JVM for the journal is 0. In the generalized equation above, the intercept is represented by $X$; in our SNIP-based regression, we calculate this value to be $1148.
* **Coefficient** is the incremental additional cost that we assume the journal charges for an increase of one point in JVM. In the generalized equation above, the coefficient is represented by $Y$; in our SNIP-based regression, we calculate this value to be $709.
* **MaxJVM** is the highest JVM value to which this equation is applied. For an equation calculated by linear regression, this would be the highest JVM in the underlying data set; higher values are outside the range of the data set and can therefore not be accurately predicted through this regression. In the generalized equation above, MaxJVM is represented by $Z$; in our SNIP-based regression, we calculate this value to be 3.207.
* **MaxAPC** is the APC assigned to any journal with a JVM above the **MaxJVM** value; because the regression cannot predict an APC in this JVM range, we must make a reasonable, educated choice. In our analysis, we chose $5000, which is approximately the highest APC currently observed in the marketplace.

This equation can either be applied uniformly across all subjects, or individually by subject (the method of applying the equation is chosen from the drop-down cell in C2). Changing the subject-specific equation parameters will only change the APC assignment for articles/journals with that subject, and only when “By subject” is selected in C2.

APCs are applied to the data set in real time; a histogram showing the distribution of APCs for the institution is shown to the right of the parameter selection table. Lines in the dataset with a JVM of “NA” are assigned the average APC across their data source. That is, APCs are calculated for all data lines with a JVM given, then are averaged by source (Web of Science or Scopus); that average is applied to all lines without a JVM given.

### Funding Allocation Parameters

This parameter sheet defines, for each paper, how the APC is allocated among the three main stakeholders: the library, granting agencies, and other discretionary funds. In practice at a particular university, these parameters will be institution-specific, chosen based on available funding and desired incentives (e.g. does the institution want to offer a greater library subsidy to incentivize publishing in higher-quality journals , but at greater cost? Or does the institution want to require authors to rely more heavily on grants, lowering costs but providing less support to researchers?).

These allocations are set in a step-wise manner: authors utilize funding from one step, and then move on to the next step if there are any remaining costs. In this way, institutions can decide how to direct their authors to obtain funding for their publications. The selected funding source will pay the entire cost, or a percentage of the entire cost, up to a given threshold. Grant funding, when selected, is always applied only to the papers in the publication data set which acknowledge a grant (the counts in the ExpectedGrantDocs column). The columns at right tell the user how much of the total cost of publishing was allocated by that step, and how much remains to be allocated in future steps; this can aid in choosing a funding strategy.

Every funding strategy should use “Any discretionary funds” or “Other author funding” with a high maximum as its final step; this ensures that after all other funding sources are exhausted, authors are expected to find their own funding from any source to pay the rest of the APC. This also preserves the basic foundation of the model, which is that authors are required to make economic decisions about how much to spend to publish their research

Use of this tab is best explained through a series of example strategies, several of which match some of the examples discussed in the PIF report (using UC Davis data as an example):

##### Strategy A: Fixed Library Subsidy (Example III from the PIF report, p.111):

Library pays a subsidy of up to $1,557 on every paper. For any costs above this subsidy, authors are required to use any other funds available to them (we assume they use grant funds if available and other discretionary funds if not).

This is a two-step example, modeled as: 

First, we apply the library subsidy to every paper: we choose **Basic** **Library Subsidy** as the funding source; we set the percentage to **100%** to signify that the subsidy covers the full cost (up to the selected maximum); and we set the maximum to the chosen value of **1557**.

Second (and last), we allocate the rest of the cost to a source of the author’s choosing: we choose **Any discretionary funds** as the funding source; we set the percentage to **100%** to signify that the discretionary funds are covering the entire remaining cost, and we set the maximum to **10000** (which effectively means there is no maximum).

As we can see in the PIF report (and in the Model Results tab, discussed below), the total cost of publishing for UC Davis is $7.49 million. In this example, the first step allocates $5.53 million to the library, and the second step allocates the remaining $1.96 million to grant funds (where available) and other discretionary funds.

##### Strategy B: Grant Funds Expended First (Example I from the PIF report, p.109):

Authors who have grant funding available must use those funds to cover their APCs. For authors without grant funds available, the library pays a subsidy of up to $1,119. For any costs above the subsidy, authors are required to use other discretionary funds available to them. This is a three-step example, modeled as:



First, we require that authors use their grant funding where available: we choose **Grant funding** as the funding source; we set the percentage to **100%** to signify that the grant covers the full cost; and we set the maximum to **10000**, signifying that the grant pays the entire APC no matter how much it costs, when there is a grant acknowledged on the paper.

Second, we apply a library subsidy to any papers which were not covered by grants: we choose **Basic** **Library Subsidy** as the funding source; we set the percentage to **100%** to signify that the subsidy covers the full cost (up to the selected maximum); and we set the maximum to the chosen value of **1119**.

Third, we allocate the rest of the cost to other author-discretionary funding: we choose **Other author funding** as the funding source; we set the percentage to **100%** to signify that the discretionary funds are covering the entire remaining cost, and we set the maximum to **10000**. Note that because all papers with grant funds were fully paid for in step 1, this step is equivalent to choosing **Any discretionary funds** as the funding source.

##### Strategy C: Partial Library Subsidy (Example IV from the PIF report, p.112):

The library pays a subsidy on every paper equal to half of the APC of the paper. For the other half of the APC, authors are required to use any other funds available to them (again, we assume they use grant funds if available and other discretionary funds if not). This is a two-step process, slightly modifying Strategy A:



First, we apply the library subsidy to every paper: we choose **Basic** **Library Subsidy** as the funding source; we set the percentage to **50%** to signify that the subsidy covers exactly half of the APC; and we set the maximum to **10000** to signify that the library covers half of the APC no matter the cost. Note that, as expected, this allocates exactly half of the cost of publishing to the library through step 1, and the remainder to other sources through step 2.

Second, as we did previously, we allocate the rest of the cost to a source of the author’s choosing.

##### Strategy D: Complex Library Subsidy (Similar to Example V from PIF Report, p.113)

In this example, we apply a variable library subsidy, based on whether the author has grant funding available and based on discipline. For any costs above this variable subsidy, authors are required to use any other funds available to them. This strategy is applied in two parts. The first part is on the Model Parameters tab:



First, rather than entering parameters, we choose a funding source of **Complex Library Subsidy**; the complex subsidy allows for varying the library subsidy based on subject and/or grant availability. We then choose Subsidy Structure # **1** to denote the proper subsidy structure to use (see below). Then, we allocate the remaining cost to other sources as we have before.

The second part, where we define the complex subsidy structure, is on the Library Subsidy Complex tab. This sheet contains three Subsidy Structure tables, numbered 1 through 3 to correspond with the structure number chosen on the Funding Allocation Parameters tab. In this way, multiple complex subsidy structures can be saved for easier switching, or even for applying multiple complex subsidies to the same allocation strategy.[[1]](#footnote-1)



Here, we edit Subsidy Structure 1, corresponding to the number we designated on the Funding Allocation Parameters tab. The table works in the same way as the Funding Allocation Parameters tab, but instead of entering a single percentage and maximum subsidy for all papers, we can specify different percentages and maximum subsidies depending on the subject and on the presence of a grant on the paper. For this strategy, we set a percentage of **100%** in every case to signify that the library pays the entire APC up to the selected maximum. However, we pick the maximum in a more granular manner: for papers without a grant, the max library subsidy is set at the baseline publishing cost of **1557**, while for papers with a grant, the max library subsidy is set at the break-even level of **1119**; this selection offers more support to authors who do not have grant funding available to them. Additionally, for the disciplines Arts and Humanities and Social Science, we have set the max subsidy for papers without grant funding to **10000**, signifying that the library pays the entire APC for papers in those fields when grant funding is unavailable to the author.

### Advanced parameters

The Advanced Parameters sheet allows the user to control two environmental parameters:

##### Grant funding percentage

In general, when publication data is imported into the Raw Data tab, it will inherently include data relating to which (or what percentage of) papers are the result of grant-funded research. However, if it is expected that actual grant funding for a particular discipline is higher or lower than the actual values in the publication output data, then those expected percentages can be applied here. Additionally, these parameters can be altered for sensitivity analysis, that is, for assessing how much a change in this environmental parameter would affect the overall outcome of the model.

To use custom discipline-specific grant percentages, change cell C3 to “y”, and enter the desired grant percentage for each discipline below in column C. Actual grant percentages from the raw publication output data are calculated for informational purposes, and are displayed in column D.

##### Percentage of articles with APCs

This factor is equivalent to **PA** from the Total Cost Equation defined in the PIF report. For each discipline, we can select what percentage of papers we expect will be published under an APC-funded model. Selecting a value of less than 100% will apply that percentage to ExpectedDocs and ExpectedGrantDocs for every line in the publication output dataset with that discipline. In this way, the percentage can also be interpreted as probability: we don’t know exactly which papers will be published under an APC-funded model and which will not, but every paper within the chosen discipline has the given percentage chance to be funded by APCs in the applied model.

Note that the scope of the PIF study assessed a world in which all publications are gold OA; as such for all of our presented analyses we use 100% for this field. However, we recognize that in a real-world application, not all journals may participate in this model. Note also that changing this parameter does NOT change the library budget available for redirection; this would need to be taken into consideration separately.

### Growth parameters

The growth parameters tab allows for predictions for a specific year in the future, based on the existing data entered. These parameters equate to the green factors in the Total Cost Equation, **APCI**, **AG**, and **y**.

Cell C3, labeled “Years forward to predict”, contains the **y** variable. That is, if the raw publication data entered on the first tab is from the year 2013, and we want to use this data to predict the costs and allocations among stakeholders in the year 2020, we would set this cell to **7**. The MCT would then use these growth parameters to project all publication output data, predicted APCs, library budget data, and extramural research expenditures forward seven years before performing any calculations related to allocating costs among the stakeholders. If we want to make assessments for the same year as the data is from, we would set this cell to **0**, and no growth parameters would be applied to the raw data; this selection was our strategy throughout the PIF report.

Column C contains the publication growth factor, **AG**. This is the yearly increase in publication volume by discipline, applied to the publication data based on the number of years forward selected in cell C3.

Column D contains the APC inflation factor, **APCI**. This is the yearly increase in publication volume by discipline, applied to the calculated APC. Note that the APC is calculated first based on the SNIP and the parameters selected in the APC Parameters tab, and then the growth factor is applied when needed.

Cells D22 and D23 represent yearly library budget growth and extramural research expenditure growth. These percentages are applied to the data points from the Model Data tab to allow for sustainability comparisons for a year in the future.

## Model Results

After setting all parameters as desired, users view the resulting funding allocations and other model sustainability measures in the Model Results tab. There is no data entry on this tab; it is purely to display results. This tab has three sections:

##### Model Summary

This section reminds the user of the institution they are analyzing and the year of the data they entered. It also displays the year which these results describe (that is, how many years forward the analysis propagates based on the data in the Growth Parameters tab). For example, if the raw data is from 2013 and the “Years forward to project” value in the Growth Parameters tab is set to 7, this will denote that the model results are for the year 2020.

##### Funding allocations

This section is the main component of the model results. It displays, by discipline and for the entire institution, three types of data:

* Summary data, including the number of total papers and the number of grant-funded papers, the total cost of publishing those papers (based on the APC projections), and the average APC.
* Funding allocation data, including the total cost allocated to the library through subsidies, to grant agencies for papers where a grant is available, and to other discretionary sources the author has access to. Additionally, the total institutional cost is presented; this is the sum of the library allocation and the other discretionary fund allocation, and it represents the amount of internal institutional money needed for publication under this model.
* Paper funding methods, listing the number of papers that are funded by each stakeholder or make use of funding from multiple stakeholders. In this way we can observe how many papers, for example, are totally covered by the library subsidy, how many use a combination of library funding and grant funding, and how many use a combination of library funding and other discretionary funding.

##### Sustainability measures

These data points represent some of the potential measures by which we can assess the sustainability for a given financial model, including break-even level calculations, comparisons of allocated funding to available funding (both internal and external), etc. Many of these measures are discussed in the PIF report.

1. An example of where this could be useful would be:

First, the library pays a subsidy of $1000 for papers without grants and $500 for papers with grants.

Second, grants pay up to $1000 when they are available.

Third, the library pays up to another $500 for papers with grants (and none for papers without)

Fourth, the author finds funding for the remainder of the APC.

In this example, the library essentially pays up to $1000 on every paper, but the author must spend some of their grant funds to be awarded the full subsidy. [↑](#footnote-ref-1)