

Mellon Foundation

Pay It Forward

Investigating a Sustainable Model of Open Access Article Processing Charges
for Large North American Research Institutions

University of California Libraries
6-30-2016

Contents

I.	Executive Summary.....	5
a.	Research Approach	5
b.	Paying for Open Access.....	6
c.	Findings	7
II.	Acknowledgements.....	9
III.	Project Overview.....	10
a.	Qualitative Data Components of the Model	12
	Author Focus Groups and Surveys.....	12
	Publisher Survey.....	12
b.	Quantitative Data Components of the Model.....	13
	Publishing Output Data.....	13
	Library Expenditure Data	13
	Research Funding Expenditure Data.....	14
	APC Data	14
	Cost-Per-Article Analysis.....	15
c.	Financial Model.....	15
	Total Cost to Institution	15
	Library APC Break-Even Point	15
	Financial Model Description	16
	Model Implementation.....	17
	Model Viability.....	18
IV.	Qualitative Data Components of the Model	19
a.	Author Focus Groups and Surveys.....	19
	Rationale	19
	Author Focus Group Methodology	19
	Author Focus Group Findings	21
	Author Survey Methodology.....	23
	Author Survey Findings.....	24
	Author Focus Group and Survey Ramifications for the Model.....	30

b.	Publisher Survey.....	31
	Rationale	31
	Methodology.....	31
	Demographics	32
	Publisher Survey Findings	33
	Ramifications for the Model	36
V.	Quantitative Data Components of the Model.....	38
a.	Publishing Output Data.....	38
	Rationale	38
	Methodology.....	38
	Scope of the Data.....	39
	Data Gathering.....	40
	Defining Disciplinary Classifications.....	43
	Defining Institutional Affiliations	44
	Data Analysis.....	45
b.	Library Expenditure Data	52
	Rationale	52
	Methodology.....	53
	Scope of the Data.....	53
	Data Gathering.....	54
	Data Analysis.....	55
c.	Research Funding Expenditure Data.....	61
	Rationale	61
	Data Gathering.....	62
	Data Analysis.....	62
d.	APC Data	69
	Rationale	69
	Methodology.....	69
	Data Gathering: APC Pricing Data	69
	Data Gathering: APC Payment Data	71

Data Analysis.....	74
Additional Analyses.....	79
Ramifications for the Model.....	82
e. Cost-Per-Article Analysis.....	83
Rationale.....	83
Methodology.....	83
Ramifications for the Model.....	88
VI. Financial Model.....	91
a. Total Cost Equation.....	91
Variable Definitions.....	91
Summary and Applications.....	94
b. Library APC Break-Even Point.....	95
Description.....	95
Library APC Break-Even Point Discussion.....	98
c. Financial Model.....	99
Model Rationale.....	99
Methodology.....	100
Economic Analysis of Journal Pricing.....	100
Estimating APC Pricing.....	102
Applying APC Pricing to Institutional Output.....	103
Library APC Subsidy.....	104
Calculating Cost Allocations.....	106
Implementation Examples.....	107
d. Disciplinary Distinctions.....	114
e. Grant Fund Tracking.....	116
f. Model Viability.....	116
Factoring in Growth.....	123
VII. Final Conclusions and Implications.....	126
a. Implications for Long-Term Costs.....	126
Overarching Factors Affecting APC Levels.....	126

Current Publishers and New Platforms	126
Author Price Sensitivity and Values	127
Library-Publisher Relations	127
b. Issues Outside of Scope	129
Author Behavior and Academic Reward Systems.....	129
Implications for Non-North American Research Institutions	130
Implications for Other Types of Library Information Resources	130
Transition to a New APC-Funded Journal System, Including Payment Mechanics	131
Appendix A: Author Focus Group Instrument	132
Appendix B: Author Survey Instrument.....	134
Appendix C: ALPSP Member Survey Instrument	148
Appendix D: Pay It Forward (PIF) Subject Mapping.....	160
Appendix E: Pay It Forward Subject Characteristics	162
Appendix F: Bibliometric Data Report Templates	164
Appendix G: Library Data Request Instructions.....	176
Appendix H: Discussion of Antitrust Potential in Large Publisher APC Agreements.....	177
Appendix I: Brief Biosketches of Project Principals	180
Appendix J: Data Sharing Plan	183

I. Executive Summary

As the global open access movement continues to grow and evolve, the question of whether a wholesale shift of the scholarly journal publishing system to “gold” open access is a viable way forward is of increasing interest. In such a shift, all journal publishers would make all scholarly articles freely available to readers, with authors or their institutions paying to publish their work when required by the publisher, rather than readers paying to read it. Lending momentum to this discussion is the fact that gold open access journals have steadily gained market share, doubling in article volume every four years and now in excess of 14% of the total journal output¹.

While gold open access doesn’t require any particular funding model, a common one is an article processing charge paid by authors, or another entity on their behalf, to cover the cost of publishing an article that has been accepted for publication. If that business model is adopted by a majority of journal publishers in the future, there are significant financial implications for the academy. As we consider the trade-offs of the status quo and various methods of achieving broad open access, questions pertaining to the long-term financial sustainability of the article processing charge business model must be carefully contemplated.

A major study conducted by the University of California, Davis, and the California Digital Library, on behalf of the University of California Libraries, and with collaborating libraries at Harvard University, Ohio State University, and the University of British Columbia addressed the financial ramifications for the types of research institutions whose affiliated scholars generate a preponderance of the scholarly literature. The project focused on large, research-intensive universities in North America and defined sustainability as costing those institutions roughly no more than, and ideally considerably less than, current journal subscription costs for comparable journals today, with a rate of growth that will be possible for these institutions to support over time. The project sheds new light on the financial viability of the article processing charge business model to create open access at a much larger scale.

a. Research Approach

Our investigation included qualitative research -- focus groups and surveys -- to learn what faculty, postdocs, and graduate students from all disciplines think about their publishing activities, open access in general, and different financial models for publishing. It also collected quantitative data about library journal budgets and institutional publishing patterns over the

¹ Archambault, Éric, Didier Amyot, Philippe Deschamps, Aurore Nicol, Françoise Provencher, Lise Rebout, and Guillaume Roberge. “Proportion of Open Access Papers Published in Peer-Reviewed Journals at the European and World Levels—1996–2013.” European Commission, n.d. http://science-metrix.com/files/science-metrix/publications/d_1.8_sm_ec_dg-rtd_proportion_oa_1996-2013_v11p.pdf.

five-year period of 2009-2013. Finally, both qualitative and quantitative information were collected from a wide range of publishers, including current APC levels, journal publication costs, and perspectives on the future of open access.

With that information, key metrics and scenarios were identified for an entirely article processing charge-funded journal publishing system to understand the financial implications and possible approaches to its sustainability, described as a financial “model”. The project’s findings may inform but do not predict whether, when, or how an entirely article fee-based journal system might emerge, nor do we intend its findings to recommend a shift to such an open access journal system.

Key to the analysis was estimating future article processing charges, if all publishers convert to that business model from their current subscription revenue. To estimate future charges, an economic analysis was done of current open access publishers and estimated charges were then used to estimate overall institutional costs, based on actual publishing patterns.

Journal article processing charges were estimated across all disciplines, based on the journal’s Source Normalized Impact per Paper (SNIP) value, to determine how much funding a given institution would need, given its current publishing patterns. Examples describe how library and grant funds might contribute to those costs and how the remaining funding gap, if any, might best be covered by new, author-controlled research funds.

b. Paying for Open Access

Using publication data from Thomson Reuters’ Web of Science and Elsevier’s Scopus databases, the “break-even” article processing charge for a library was calculated (very roughly, the library’s journals subscription expenditures divided by the number of articles published by corresponding authors from that institution). With that baseline data point, we developed examples of how institutions might organize funding for article processing charges incurred by their authors.

If the only source of funds for article processing charges is the library’s journal budget, given estimated future charges, institutional research publishing productivity will determine whether the library journal budget can cover publication fees. Less research-intensive institutions could entirely fund their authors’ publishing costs from the library’s budget, while more research-intensive institutions would likely have a significant funding gap.

However, library budgets are not the only possible source of funds to pay article fees. Other possible sources include grant funds, and these are an important factor in our analysis of financial sustainability beyond what library budgets might support. A flexible funding model is necessary to allow research-intensive institutions to combine funding sources to cover article

processing charges, while containing costs over time. Such a model is based on a few key assumptions:

- Much research at large North American research institutions is grant funded, and those grants are a possible source of funding to pay article processing charges (as is the case today);
- In a fully article processing charge-funded journal market, article processing charges may increase rapidly unless there are controls to limit that market behavior;
- To ensure that article processing charges remain affordable in the future, it is essential to introduce competition for authors into the journal market. That can be accomplished by giving authors some financial responsibility in deciding where to publish, e.g., using funds that they control directly;
- Over time, article processing charge levels will be differentiated according to authors' perceived "value" of publishing in a particular journal. A proxy for that value is Impact Factor (IF) or Source Normalized Impact per Paper (SNIP), such that article processing charges will predictably correlate with IFs and SNIPs.

A sustainable funding model allows for various scenarios that combine funding sources in different ways, including:

- Library journal budgets, redirected to provide some level of APC subsidy;
- Grant funds that are available to some authors;
- Discretionary research funds available to all authors, under various conditions.

c. Findings

Three major conclusions from the project are as follows:

1. For the most research-intensive North American research institutions, the total cost to publish in a fully article processing charge-funded journal market will exceed current library journal budgets;
2. This cost difference could be covered by grant funds, already a major source of funding for publishing fees; but
3. Ultimately, author-controlled discretionary funds that incentivize authors to act as informed consumers of publishing services are necessary to introduce both real competition and pricing pressures into the journal publishing system. Discretionary funds for authors exist today, in the form of research grants, personal research accounts, endowed chair funds, and departmental funds, but the consistent application of these funds for this purpose would, in some cases, require new funding from the institution.

A few notes about the financial model provide important context. First, the model accounts for a financial surplus (or profit) for publishers to support ongoing innovation, but does not attempt to replicate current profit margins of journal publishers, whether commercial or non-profit (e.g. scholarly societies). Second, the information available on current article processing charges is derived almost entirely from STEM fields, which historically have higher subscription costs than social science and humanities fields. Since the article processing charge estimations

are based on available data, they are likely overestimates for non-STEM fields and different formulae will be needed in other disciplines as more data become available.

A wholesale shift to a new business model for publishing, in which authors or their institutions pay at the point of publication, instead of institutions paying to subscribe, would have far-reaching implications for the distribution of costs for scholarly publication, the reach of those publications, equity in the ability to publish research, incentives for innovation in online publishing technology, and much more. This project attempts to shed some light on one important aspect of the shift: whether it would be financially sustainable for the large, research-intensive institutions in North America that currently publish the largest share of the world's research. The answer is complex but points a way toward a possible funding model that could introduce the right incentives to both manage costs and improve the system over time, while achieving the benefit of fully open access to research.

II. Acknowledgements

The University of California 'Pay it Forward' project team thanks the Andrew W. Mellon Foundation for its generous support of this project and its ongoing commitment to improving scholarly communication in the digital age.

III. Project Overview

In 2014, with support from the Andrew W. Mellon Foundation, the UC Davis University Library and the California Digital Library collaborated on a planning project to investigate the economic implications of the “article processing charge” (APC) funding model for open access to scholarly journals. The APC model is a popular variant of the “gold” open access model in which publishers are compensated for their effort by authors (or their proxies) at the point of publication rather than by charging subscription fees for access to the journals. The “gold” open access model also encompasses journals with other sources of funding so that authors are not charged. UC’s interest in this issue is propelled by the reality that researchers at the University of California author a substantial proportion of the scholarly literature (over 2% of all papers in Web of Science and Scopus have a UC author) and are strong supporters of open access (as evidenced by the UC Academic Senate’s 2013 open access policy), but converting the cost of scholarly communication to an “author pays” or, potentially, an “institution pays” model has huge implications for large research institutions that generate a disproportionate amount of the literature. Finding the right financial model to pay for a more open form of scholarly communication, one that does not replace the imperfections of today’s publishing economics with a new set of structural deficiencies, requires significantly more evaluation.

The planning process included a proposed methodology for a large-scale investigation of this topic. The Mellon Foundation subsequently funded the proposal, and the project was conducted throughout 2015 and into early 2016. Along with UC Davis, the California Digital Library, and the nine other UC campuses, the project included the participation of three major research institutions- Harvard University, The Ohio State University, and the University of British Columbia. This focus on large, North American research universities recognizes that a relatively small number of North American research universities generate a substantial percentage of published research papers, and that North American universities and their research are funded differently than in other parts of the world. We sought out partner institutions that were both large (i.e., Carnegie RU/VH) research institutions with heterogeneous research programs that include humanities, social, life and physical sciences, and that had extensive experience with open access publishing (e.g., a faculty mandate or extensive library engagement). The inclusion of these partner institutions helped to validate and provide a basis of comparison for the data gathered at UC.

The key issue this project aimed to address is whether a large-scale conversion to open access scholarly journal publishing funded via APCs would be viable and financially sustainable for this class of large North American research-intensive institutions. Sustainability in this context is defined as costing these institutions roughly no more than, and ideally considerably less than, current journal subscription costs for comparable content today, with a rate of growth that will be possible for these institutions to support over time. We consider “viability” to further encompass the willingness of authors to publish under such a model and the likelihood that

they will do so if the option is available and sufficiently congenial to them, as well as the additional likelihood that a scholarly publishing infrastructure optimized for open access will allow research and scholarship to maintain its present degree of quality.

In order to properly investigate these notions of sustainability and viability, the project team gathered a wealth of qualitative research, primarily in the form of focus groups and surveys. This helped to surface what faculty, postdocs, and graduate students from a range of disciplines think about their publishing activities, open access in general, and different financial models for publishing. We also collected considerable quantitative data, including library journal expenditures and institutional publishing patterns over the five year period of 2009-2013. Finally, we collected both qualitative and quantitative information from a wide range of publishers, including current APC levels, journal publication costs, and perspectives on the future of open access.

Compiling useful, accurate, and comprehensive data across the complicated range of topics outlined above required the expertise of a diverse group of project participants. Key personnel participating in this project (brief biosketches are provided in Appendix I) include MacKenzie Smith (University Librarian, UC Davis University Library), Laine Farley (Executive Director Emeritus, California Digital Library), Ivy Anderson (then-Interim Executive Director and Director of Collections, California Digital Library), Greg Tananbaum (Consultant, ScholarNext), Mathew Willmott (Project Analyst, California Digital Library) Professor David Solomon (Michigan State University), Professor Bo-Christer Björk (Hanken School of Economics, Helsinki, Finland), and Professor Mark McCabe (University of Michigan and Boston University). Collectively, this group comprises the “modelling team” that is referenced in various sections of this report. We were additionally joined by Dr. Carol Tenopir (University of Tennessee, Knoxville) for an in-depth qualitative analysis of authors’ attitudes towards open access. Finally, we collaborated with the Association of Learned and Professional Society Publishers (ALPSP) to engage their member publishers, and the information companies Thomson Reuters (Web of Science) and Elsevier (Scopus) for their bibliographic database coverage of authorship patterns across the academic disciplines. The specific contributions of each project participant and partner are called out in the body of the report.

Our research clearly illustrates where there are gaps in our access to important data, or a lack of consistent data practices across libraries, universities, and publishers that make this type of analysis difficult. We believe that the findings are defensible, interesting, and useful, but are the first to admit that much more needs to be done.

To assist readers with the overall flow of the report, we have provided a concise summary of each section below. This is designed to provide a high-level overview of the work that we undertook, the data we gathered, and how these data informed our model. Fuller descriptions can be found in the detailed sections of the report.

a. Qualitative Data Components of the Model

Author Focus Groups and Surveys

To better understand researcher attitudes toward the current publishing system and its potential alternatives, the project team undertook both a series of focus groups and a survey of authors. These activities were designed to establish a better understanding of what scholars think about research accessibility, current publishing practices, gold open access and article processing charges, and the future of scholarly publishing and its economic models. Ten focus groups were conducted across five campuses, encompassing 77 participants (46 faculty members and 31 graduate students). The survey was distributed to a total of approximately 15,000 academics across four campuses, with an overall response rate of 14.1%. Both the focus groups and the survey examined the extent to which the research community would accept changes to the scholarly communication model, and whether this audience could identify and embrace possible benefits of such a shift. These attitudes were probed from both the consumption (i.e., the reader) and the production (i.e., the author) perspectives. With respect to the former, respondents simultaneously acknowledged both their personal satisfactions with research accessibility at their institutions and their awareness that the wider community would benefit from increased open access. From the authoring point of view, concern was expressed about the financial ramifications of widespread open access, not only personally, but also as it might impact departments, disciplines, universities, and developing countries. When asked about the importance of various journal criteria in deciding where to publish, authors placed journal quality above all else (followed closely by journal fit and audience), while Open Access ranked as least important. These findings suggest that it will be very difficult, impossible in some cases, to convince authors to switch to new journals in a new publishing system purely for the sake of open access. Additionally, there remains a lingering notion that APCs are akin to “pay to play.” Authoring attitudes varied across subject areas, with disciplines in which open access is more common and more external grant money is available generally more supportive of OA. To the extent that the publishing environment trends towards gold open access, respondents favor an approach that does not place undue pressure on authors to identify APC funding sources, that is easy to implement, and that is perceived as equitable both within and across institutions. Participants also demonstrated a preference for the library to play an active coordinating role in any transition to the open access model. Survey results clearly demonstrated author price sensitivity, in terms of how much they were willing to pay from different funding sources. There was also some sentiment for strengthening market principles in the gold OA publishing environment, consistent with the modeling developed over the course of this project.

Publisher Survey

The project team worked with the Association of Learned and Professional Society Publishers (ALPSP) in the development and execution of a member survey. The aim was to better understand publisher opinions and behaviors pertaining to open access. This encompassed not

only current publishing activities, but also the extent to which publishers are planning for a continued growth in gold open access. All 240 ALPSP publisher members received the survey, and 30% responded. The vast majority of participants (86%) were some form of not-for-profit publisher. The ALPSP survey results indicate that open access is only modestly impacting both current publisher operations and R&D planning. This appears to be largely due to the modest effect OA has had on publication volume, overall revenues, and relationships with libraries to date. It is possible that these attitudes are driven by the relatively modest open access output among survey respondents, which, at 9%, is below the overall industry average. Looking toward the horizon, publishers are planning to grow both the number of gold open access journals and hybrid articles they publish, with a perceived need to adjust operations and workflows in response as well. Nearly three in four respondents do not view open access as a challenge to the future health of their organizations. Overall, the survey results suggest that publishers are not presently threatened by gold open access, seeing it as a growing part of their futures, and that they are—with some allowance made for cultural and disciplinary differences—open to a clear-eyed assessment of this project’s findings.

b. Quantitative Data Components of the Model

Publishing Output Data

We collaborated with Thomson Reuters and Elsevier to obtain several sets of bibliographic data from Web of Science and Scopus for use in our modeling efforts, encompassing journal articles, review articles, and conference proceedings published from 2009 through 2013². These datasets included article-level data from Web of Science, with standard bibliographic information and related data on authors, their affiliations, and grant acknowledgement statements; aggregated article data from both Scopus and Web of Science at the journal, year, institution, and document type levels, including publication counts and authorship patterns; and journal-level data from both Scopus and Web of Science, including bibliographic data, citation metrics, and other statistics. Data from the two sources was synthesized to generate a full picture of the publication output of our partner institutions over the course of the five-year study period. Using these datasets, we constructed a disciplinary classification scheme, selected a strategy for assigning APC payment responsibility to a particular institution, and generated lists of in-scope materials to use when gathering data from our partner institutions.

Library Expenditure Data

We gathered data from the libraries at our partner institutions relating to subscription expenditures on materials within the scope of the project (i.e., the set of journals and conference proceedings covered by Web of Science or Scopus). Data were gathered through direct collaboration with representatives at each library, and included the overall expenditure on in-scope materials either through direct, single-title subscriptions or through package

² For convenience sake, this report uses the terms to “articles” and “journals” to encompass conference papers and conference publications as well.

purchases. Access to materials obtained through aggregators or other third-party providers were considered out-of-scope for the project, as these do not constitute direct agreements with the publishers where funding would be easily redirected from subscriptions to APCs.

The data obtained from libraries were aggregated by year for the period of 2009 through 2013. Data were also broken down into expenditures by format (print, electronic, or combined) and by whether the pricing was negotiated by a consortium. Additionally, data on expenditures for memberships with gold open access publishers were requested, since these are funds that may be redirected towards APC payments in a new model. Based on these data, total expenditures were calculated that we consider eligible for redirection to APCs in a fully gold open access environment.

Research Funding Expenditure Data

Research expenditure data for each partner institution were gathered from publicly available sources such as the National Science Foundation's Higher Education Research and Development (HERD) survey, in order to determine the extent to which APC payments would consume overall grant funding. These data were also compared to the publication output data gathered previously, to observe how the disciplinary distribution of grant funding compares to publication output. Finally, grant agency policies were investigated and compared to the grant acknowledgment data gathered as a part of the publication output data, to affirm our assumption that the majority of papers with a grant acknowledgment statement could be charged to those grants.

APC Data

Various types of APC data were gathered for a thorough analysis of both publishers and authors in setting and paying APCs. List price APC data for full OA journals were gathered from a longitudinal study led by Heather Morrison and were updated by our own investigations. These prices were shown to correlate with citation metrics such as SNIP and Impact Factor; however, disciplinary differences were not significant enough to draw any conclusions due to the paucity of gold OA journals in non-STEM fields. We mapped the pricing dataset to our publication output data set, to estimate how much researchers at our partner institutions paid in APC charges for publications in existing full OA journals over the course of the study (~\$1,892), as well as the average APC set by publishers for journals in which authors at our partner institutions published (~\$1,864).

Additionally, we gathered data from various European databases recording actual APC payments made by granting agencies or institutions on behalf of authors. These datasets showed current behavior around OA publication, including the observation of a much more mature market for OA publication in some disciplines (life sciences, clinical medicine) than others (arts and humanities, business and economics, mathematics). A significant difference was also observed between the price paid for publication in a full OA journal (average \$1,907) and a hybrid journal (average \$3,045). The averages calculated here serve as valuable points of

comparison for our financial model; additionally, components of the data were used in the “Cost-Per-Article Analysis” section, as well as in predicting future APC pricing in the “Financial Model” section.

Cost-Per-Article Analysis

In the “Cost-Per-Article Analysis” section, we attempted to ascertain what a sustainable journal publishing operation might cost on a per-article basis. We first explored the possibility of constructing a ground-up cost model. This was ultimately dismissed as unfeasible for a variety of reasons, notably the high degree of variability in what constitutes publishing services. In its place, we examined actual cost data from a variety of sources, including tax forms, literature reviews, analysis of gold OA journals in which our authors publish, and discussions with publishers. This process allowed us to develop a floor and average Cost-Per-Article, including a 13% surplus to fund ongoing innovation. This sustainability range, from \$1,103 at the low end to \$2,566 at the high end, helped to establish the viability of the financial model we developed, and test whether it could provide sufficient income for publishers to sustain their core functions.

c. Financial Model

Total Cost to Institution

We developed an equation to calculate the total financial responsibility that an institution would bear under a fully APC-funded open access system. This represents the amount of funding that the institution must secure, either through redirected journal subscription funds, author-applied grant funds, and/or other institutional funding. The equation contains several variables, relating to factors such as publication output, expected average APC, and growth of publications over time. Each variable is defined, specifics about the variable are discussed, and approaches to assigning value(s) to the variables, based on the data gathering activities described in the “Quantitative Data Components of the Model” section, are explained.

Additionally in this section, we present and discuss the criteria used to develop the financial model for how an institution might organize funding and allocate costs, within the structure of the cost equation. These criteria are based on economic theory and the conclusions drawn from our author focus groups and survey. The five criteria identified are that libraries should continue to play a major funding role in any scenario; grant funding should be considered a legitimate and routine source of funding for open access publication charges; establishing the right marketplace incentives should be a key component of any funding model; to achieve a functional incentive structure, authors should have “some skin in the game”; and authors should not bear an undue burden in an APC-driven model.

Library APC Break-Even Point

We identify the price points at which each library would be able to cover the entirety of publication charges for their researchers, given only the funds identified as redirectable

expenditures in the “Library Expenditure Data” section. These calculations provide a price level that we know the institution can support, which we further use as we define and implement our proposed model. For each partner institution, this break-even point is calculated as the total redirectable expenditures in a given year, divided by the publication output of the institution in that year. In general, high-output research institutions have break-even points well below some of the averages that we see in our APC and Cost-Per-Article data gathering work, implying that their journals budgets would be unable to cover publication at the cost that we’re seeing today. Institutions with a more modest research output have higher break-even points, implying that they may be able to support their researchers’ publishing behaviors at current price levels.

We repeat the calculation while removing all papers that acknowledge a grant. This approach essentially assumes that grant agencies will cover publication charges for papers resulting from their research funding, and then calculates the library’s break-even cost for the remaining papers. As discussed in the “Research Funding Expenditure Data” section, a large percentage of our partner institutions’ grant funding in the US (~72%) comes from federal agencies whose policies treat publication costs as an allowable expense, and many private funders have adopted such policies as well. As expected, break-even levels are much higher in this approach and were in general found to be within the library’s budget envelope for even those institutions with the highest research productivity, because libraries have fewer papers to cover with their redirected expenditures.

We also discuss potential variability in the data as a result of several factors inherent in the data themselves or our interpretation of them. In general throughout our analysis we select the worst-case scenario, such that the break-even levels we present are the lowest that we could reasonably expect to see from each institution. However, there are factors which could potentially raise the break-even levels observed to more optimistic levels. Additionally, we discuss how we might be able to apply discipline-specific characteristics to the data to create a more accurate picture; however, in our project this was either not possible or was imperfect and subject to variability, mainly due to challenges in the available data (such as its heavy dependence on data from scientific and medical fields rather than social sciences and humanities).

Financial Model Description

We next define the structure of our proposed APC-funded journal financial model, based on the cost equation and the criteria for a viable and sustainable model. In our model, APC payments are distributed among three potential funding sources: the library (with funds redirected from journal subscriptions), research grant funds, and other author-controlled discretionary funds that might be drawn from various institutional sources. Marketplace competition is introduced into the model by offering a library subsidy up to a specific APC amount, and requiring the author to find additional funds, either from grants or other discretionary funding, to cover APCs above that price tag. This price pressure allows authors to act as thoughtful consumers, making

choices about the value of a journal as compared to the APC in light of other uses they might have for those funds and thereby controlling costs.

We undertake an economic analysis to observe the relationship between price and journal quality (as perceived by the author). To put this analysis into practical terms, we use the journal's SNIP value as a proxy for journal quality, and run a linear regression on a subset of the APC pricing data described in the "APC Data" section. The equation generated by this regression allows us to predict the APC of any journal, given that journal's SNIP value. We use this equation to predict the APC for every article in our bibliometric data set, thereby calculating the total cost of each institution's scholarly publishing activities for each year in our study.

Finally, we discuss the various strategies for selecting an appropriate library subsidy. The general goal of the subsidy is to offer significant institutional support for publication, while still requiring authors to contribute some amount of funding, especially for publication in more expensive journals (which, under our economic analysis, will tend to be of higher quality). Possible subsidies include the break-even level calculated in the previous section, as this would ensure that the library pays no more than they have been paying under the subscription model; the Cost-Per-Article, calculated previously, as this would allow the library to cover the actual costs of publishing without contributing to unreasonably high profit margins; and the predicted APC for a journal with SNIP = 1.0, as this would cover the cost of an average, baseline journal in any field, while requiring authors to pay for any additional value above and beyond that baseline.

Model Implementation

In this section, we present and discuss five example implementations of our model, and look at the actual predicted allocation of costs that these examples would generate at three of our partner institutions: one high-output research institution, one average large research institution, and one low-output institution. The first example requires authors to apply grant funds, if available, to cover the entire APC; if no grant funds are available, then the library pays a subsidy up to the calculated break-even point, and the author must find discretionary funding for any additional costs. In the second example, the library pays a subsidy up to the calculated break-even point for every paper; the author then uses grant funds or other discretionary funding to cover the remaining costs (authors are assumed to use grant funds if available). In the third example, the library pays a subsidy based on the predicted baseline journal cost, and the author then uses grant funds or other discretionary funding to cover the remaining costs. In the fourth example, the library pays exactly half of every APC, and the author uses grant funds or other discretionary funding to cover the other half of the publishing costs. In the fifth example, the library pays a subsidy based on the predicted baseline journal cost for papers where grant funding is unavailable, and a significantly reduced subsidy for papers where the author does have grant funding available; the author then uses grant funds or other discretionary funds to cover the remaining costs.

Model Viability

Many factors must be taken into account when assessing the viability of this financial model, and some implementations may be viable for certain types of institutions but not for others. For the high-output institution, for example, offering a library subsidy at the predicted baseline APC would greatly increase the cost to the institution; assigning a library subsidy at the break-even level would be much more viable given current budgets. However, for the low-output institution, the break-even level is high enough that it would fully cover most APCs, and it would thereby not properly incentivize authors to act as thoughtful consumers; the predicted baseline APC may therefore be a more appropriate subsidy.

Considerations beyond just the cost resulting from the library subsidy are discussed as well. The grant funds needed for some of the examples are compared to the total external funding available to the institution (as calculated in the “Research Funding Expenditure Data” section). For some institutions and some examples, less than half a percent of their external funding would be needed to support publication; for one or two others, close to 2% would be needed. Additionally, many authors, especially those in disciplines with lower levels of grant funding, will need additional support in the form of other discretionary funding. These funds, which could potentially come in the form of a yearly research account, must be available to the author for any purpose, not just publishing. They must also be taken into consideration when assessing the viability of the model. Finally, some of the challenges and variable components of the model are discussed, including unintentional incentives for publishers to raise APCs; disciplinary distinctions and how they could affect the model; and the application of grant funding to all publications which acknowledge a grant.

IV. Qualitative Data Components of the Model

a. Author Focus Groups and Surveys

Rationale

A socio-technical system is a social system (i.e., academic publishing) that is built upon a technical base (i.e., the technologies and technical workflows required for publication). The shift to an APC funding model implies the introduction of a new socio-technical system for scholarly publishing. This will undoubtedly produce changes in workflow for numerous stakeholders, notably the researchers themselves. Literature from business management, informatics, science and technology studies, sociology, and anthropology indicates that the successful and sustainable introduction of any new socio-technical system requires the “buy-in” of these affected stakeholders.³ Buy-in is predicated upon a stakeholder’s perception that the proposed changes required for the new system result in added “value” over that provided by the legacy system being replaced.⁴

Changes to the current publishing model will potentially impact where authors choose to publish and the visibility of their research, among other factors. This impact will be rendered even more significant under scenarios in which they, as authors, will play an active role in the allocation of funds to cover a range of scholarly activities. Establishing a baseline understanding of stakeholders’ perception of “value” and potential “buy-in” among authors to a different scholarly publishing system was therefore a necessary component of this project.

To understand researcher attitudes toward the current publishing model and its potential alternatives, the project team worked with Professor Carol Tenopir and her colleagues at the University of Tennessee to develop a series of focus groups followed by a survey of authors. The methodology and findings of these activities are discussed below.

Author Focus Group Methodology

The goal of the focus groups was to gather qualitative data that addressed opinions and comments relating to research accessibility, participants’ current publishing/authorship practices, opinions on gold open access and article processing charges, and the future of scholarly publishing and its economic models. In addition, the focus groups were used to inform development of the later survey questionnaire.

During the month of February 2015, the research team from the University of Tennessee, Knoxville developed a focus group research instrument with the input of the project team. The

³ Steiner, Gerald 2008. Supporting sustainable innovation through stakeholder management: a systems view. *International Journal of Innovation and Learning* 5(6): 595-616.

⁴ de Waal, Benny ME and Ronald Batenburg 2014. The process and structure of user participation: a BPM system implementation case study. *Business Process Management Journal* 20(1): 107-128.

questionnaire went through several rounds of edits, followed by pilot testing with faculty members and graduate students from the University of Tennessee, Knoxville, who were unaffiliated with the project and relatively unfamiliar with the topic. This further refined the questionnaire to better accommodate those who may be less familiar with the topic of open access publishing, and to promote greater depth and higher quality of responses from participants. In addition, in order to provide a useful discussion prompt for those less familiar with open access publishing and article processing charges, a chart listing a few open access journals and corresponding article processing charges (APCs) from different fields was created to give participants a frame of reference for the discussion of APCs and publishing models. The full final focus group instrument is provided in Appendix A.

For the focus groups, both faculty and graduate student researchers were recruited from a broad range of subject disciplines at each participating university. With the help of key contacts, primarily through the libraries at each participating research institution, Dr. Tenopir’s team achieved a total participation of 77 scholars in 10 groups. The group totals are as follows:

Table 1: Focus Group Participants by Institution

School	Faculty	Graduate Students	Total
University of British Columbia	9	8	17
University of California, Irvine	8	2	10
University of California, Davis	10	7	17
Harvard University	6	5	11
The Ohio State University	13	9	22
Total	46	31	77

Institutional Review Board (IRB) approval for this study was obtained by the University of Tennessee, Knoxville, Office of Research Compliance. Approval was also obtained by each of the participating institutions. Informed consent statements for both the University of Tennessee and the participating institution were read and signed by each participant upon their arrival at the focus group site. The research team provided a gift bag, which included a journal and a thumb drive, to each participant.

Participants came from a wide range of subject disciplines, including arts and humanities, physical and biological sciences, social sciences, medicine, law, and mathematics. Each session lasted between an hour and 15 minutes and an hour and a half and was audio recorded to ensure accuracy of results. The researchers held debriefing sessions after each group, and all

notes were transcribed (yielding approximately 300 single-spaced pages) for later use in analysis.

Author Focus Group Findings

Each focus group opened with a discussion of the participants' own practices as consumers of research. They discussed how they access what they need, their satisfaction with the level of access they have at their universities, what obstacles or barriers to access they face, and how they overcome these barriers. Satisfaction with level of access tends to be high, although barriers do exist. Many participants described using personal and professional networks to locate and access what they need, as well as subject-based repositories such as the arXiv, if they cannot locate publications through their library. Both faculty and graduate students were enthusiastic about the depth of their library collections and the services provided by their library, such as interlibrary loan.

Next, the discussion moved onto participants' practices as authors. Participants discussed the parameters of publishing, the venues in which they publish their work, and what makes a publication outlet appealing or not. For many, what counts as "publishing" goes beyond peer-reviewed journals. Conference proceedings, blogs, and datasets were all considered credible forms and outlets for publication depending on the respondent's discipline. However, when it comes to the criteria by which participants judge a publishing outlet, the concern for getting hired (graduate students), achieving tenure (newer faculty), and building and maintaining a reputation (more advanced faculty) means that the prestige and reputation of a publishing outlet are the most important factors they consider when deciding where to publish their own work. Peer review and high quality or highly rated journals were mentioned by many as being very important in their decision making.

For participants in the fine arts/humanities, speed of publication, fitness within discipline or topic, readership, and prestige of the journal were the most important concerns. They wanted to publish where the stars of their field were publishing, because even if they got rejected, they had great feedback to improve their work. For life scientists, impact factor, turnaround time, metrics, prestige, and open access were all deemed important. They were concerned with publishing in a journal with easily captured metrics for purposes of tenure. More than any other field participating in the focus groups, life scientists were concerned about open access. They wanted to publish in open access journals, largely because access to potential audience was equally important to them. Similarly among engineers, prestige and open access were the most important factors. They also wanted to publish where the groundbreaking work was being published. Physical scientists cared about impact factor, reputation, readership, speed of publication, open access, and number of citations. For social scientists, it was journal fit, impact factor, and robust peer review. There was a tension between open access and

reputation. They recognized the value of open access, but sometimes impact factor won out over the benefits of openness. For computer science, impact factor was the critical key in determining where to publish.

This discussion was then directed toward the importance of accessibility, open access publishing, and article processing charges (APCs). There was a clear distinction between perspectives when talking about OA from a reader's perspective and from an author's perspective. From a readership perspective, open access is widely considered ideal. Participants viewed the ability to access and read scholarly papers regardless of institutional membership as a plus, especially given lack of availability for certain titles at different institutions. However, there are issues related to the quality of open access sources, participants' ability to evaluate level of quality, and the ability to access older papers. From an authorship perspective, the picture is much more complex. Opinions about publishing in open access outlets, and the model in general, ranged from extremely positive to extremely negative, with most participants somewhere in the middle. For those in the middle, the decision on whether to consider an open access journal to publish depends on other factors, such as the prestige of the title, the importance/perceived quality of the particular paper they are submitting, and desires of co-authors or supervisors.

Reactions to a sample list of open access journals and corresponding APCs (provided by the facilitators) raised many questions, including what the fees are used for and what accounts for the variation in charges. Most felt that journals charged what they could; therefore, some of the more recognizable titles were seen as prohibitively expensive for those authors without ample grant funding. Some participants, such as those in humanities and social sciences, immediately associated APCs with the stigma of "pay to play," while others, particularly in medicine and health sciences, were more accustomed to these charges. In addition, those who practiced self-archiving (especially physical sciences and mathematics) saw no point in paying for what they were already making publicly available. Among the graduate students in particular there was little prior knowledge of, or experience with, the notion of APCs. Some raised the concern that APCs would disadvantage authors in institutions that could not assist with payment.

Because the majority of participants were in favor of the idea of open access from the perspective of readership, they were then prompted to think about how to make the model work. There was much discussion about funding models that did not put pressure on the individual author to pay for publication out of their own pocket, with experience levels ranging from graduate students who had never published to faculty with extensive editorial experience. A number of these suggestions, including the use of funds from grants, university accounts, and discretionary budgets, are contemplated within this report's model.

Author Survey Methodology

The survey questionnaire was developed to allow for more focused questions to be addressed to a broader group of participants in order to gather quantitative data that could be analyzed statistically and extrapolated to the wider populations. Topics, terminology, and scenarios that emerged from the focus group discussions were incorporated as guiding constructs in the development of survey questions. During the month of May 2015, the research team developed the survey questionnaire with input and approval from the economic modeling team and principal investigator. Multiple rounds of development and testing resulted in a questionnaire of between 20 – 30 questions (depending on skip and display logic routes of participants' answers). The full questionnaire can be found in Appendix B.

Dr. Tenopir's team sought a representative sample of academic researchers across a broad range of subject disciplines at four research universities. Respondents included faculty, graduate students, and post-doctoral researchers from the University of British Columbia, The Ohio State University, the University of California, Irvine, and the University of California, Davis. The criterion for inclusion stated that respondents must be employed as faculty or staff at the university, or be enrolled as a graduate student at the university. The survey was distributed to a total of approximately 15,000 academics. With 2,121 responses, the overall response rate was 14.1%.

As with the focus groups, IRB approval for this study was obtained by the University of Tennessee, Knoxville, Office of Research Compliance. Approval was also obtained by each of the participating institutions. An informed consent statement was included on the first page of the survey, and consent was obtained by having participants click "next" after reading the statement. Within both the recruitment email and the informed consent statement, participants were informed that as an incentive, they would have the option to be included in a prize drawing for an iPad Mini. If they chose to participate, they would be redirected at the end of the primary survey to a new, separate webpage where their email addresses would be collected.

The survey was developed and distributed using Qualtrics software. A detailed checklist was provided for the distributing partners at each participating institution. Before launching the survey, pilot test links were sent to a small sub-sample of academic researchers at the University of Tennessee, Knoxville, and the four participating universities from which subjects would be drawn for participation in the actual survey (faculty and graduate students, $n = 30$). Pilot tests were conducted both in-person, with the authors present for questions and feedback, and remotely via the test link and emailed feedback. These pilot testing procedures were used to ensure that language was clear, the sequential ordering of questions was logical, and that there were no technical problems with the survey link.

After ensuring that the survey was working properly, live links were sent to librarian distributors at the four participating research institutions. Distributors then sent the links to targeted email distribution lists with the goal of ensuring roughly equivalent sampling across different disciplines and position types (e.g., faculty, graduate students) at the university. The survey was open from May 20, 2015, to June 10, 2015 (approximately three weeks). After two weeks, a reminder email was sent by librarian distributors thanking those who had already participated and reminding others of their opportunity to participate. The final number of respondents was 2,121. After cleaning the data, which included removing all respondents that did not provide at least one independent variable response and one dependent variable response, the final $n = 2,021$.

Author Survey Findings⁵

The survey attempted to understand possible considerations that academic authors may take into account when choosing a journal outlet in which to publish their work. These eight factors, drawn from both previous research^{6,7} and the focus groups conducted among this population prior to the survey include audience, editor or editorial board, fit with scope of journal, impact factor, likelihood of acceptance, open access, quality and reputation of journal, and time from submission to publication. Respondents were also given the opportunity to list other factors they consider when choosing a journal outlet for their research.

Respondents were split roughly in half between faculty ($n = 935$) and graduate students ($n = 915$) with postdoctoral researchers/other constituting the remaining respondents ($n = 170$) (Table 2). Faculty are largely made up of Assistant ($n = 219$), Associate ($n = 218$), and full Professors ($n = 480$). The large majority of graduate students are pursuing doctoral degrees ($n = 730$).

⁵ Portions of this section draw from, "What Matters to Authors of Scholarly Articles: The Influence of Journal Factors and Potential Audience on Journal Choice", by Carol Tenopir, Elizabeth Dalton, Allison Fish, Lisa Christian, Misty Jones, and MacKenzie Smith, submitted for peer review February, 2016.

⁶ Mabe, M., & Mulligan, A. (2011). What journal authors want: Ten years of results from Elsevier's Author Feedback Programme. *New Review of Information Networking*, 16(1), 71–89. <http://doi.org/10.1080/13614576.2011.574495>

⁷ Solomon, D. J., & Björk, B.-C. (2012). Publication fees in open access publishing: Sources of funding and factors influencing choice of journal. *Journal of the American Society for Information Science & Technology*, 63(1), 98–107. <http://doi.org/10.1002/asi.21660>

Table 2. Career Status Demographics of Survey Respondents

Position	Frequency	Percentage
<i>Faculty*</i>	935	46%
Adjunct professor / lecturer	(3)	(<1%)
Assistant professor	(219)	(23%)
Associate professor	(218)	(23%)
Professor	(480)	(51%)
Research faculty	(4)	(<1%)
Other	(10)	(1%)
<i>Students</i>	915	45%
Masters students	(129)	(14%)
PhD students	(730)	(80%)
JD students	(27)	(3%)
MD students	(27)	(3%)
Other	(2)	(<1%)
<i>Postdoctoral fellows / other</i>	170	8%
Total	2,020	100%

* One faculty respondent did not provide their position title.

The largest group of participants fell into the category of life sciences/medicine (n = 623), followed by social sciences (including business, law, and education) (n = 549), arts and humanities (n = 350), engineering and computer science (n = 268), physical sciences (n = 175), and mathematics (n = 45) (Table 3). Ten respondents chose “other” with no further information specified.

Table 3: Disciplinary Demographics of Respondents

Subject Discipline	Frequency	Percentage
Arts & Humanities	350	17%
Engineering & Computer Science	268	13%
Life Sciences and Medicine	623	31%
Mathematics	45	2%
Physical Sciences	175	9%
Social Sciences (including Business, Education, & Law)	549	27%
Other	10	<1%
Total	2,020	100%

The results of this survey indicate that a journal’s reputation and fit with the author’s work remained the two most important factors that authors take into consideration, followed by the journal’s audience. Even though impact factor was considered to be important, it trails behind ensuring that the author’s work is situated alongside other reputable research and is being read by the intended, or “right” audience. Taken together, it is evident that reputation building within a specific field is at the heart of what matters most to academic scholars.

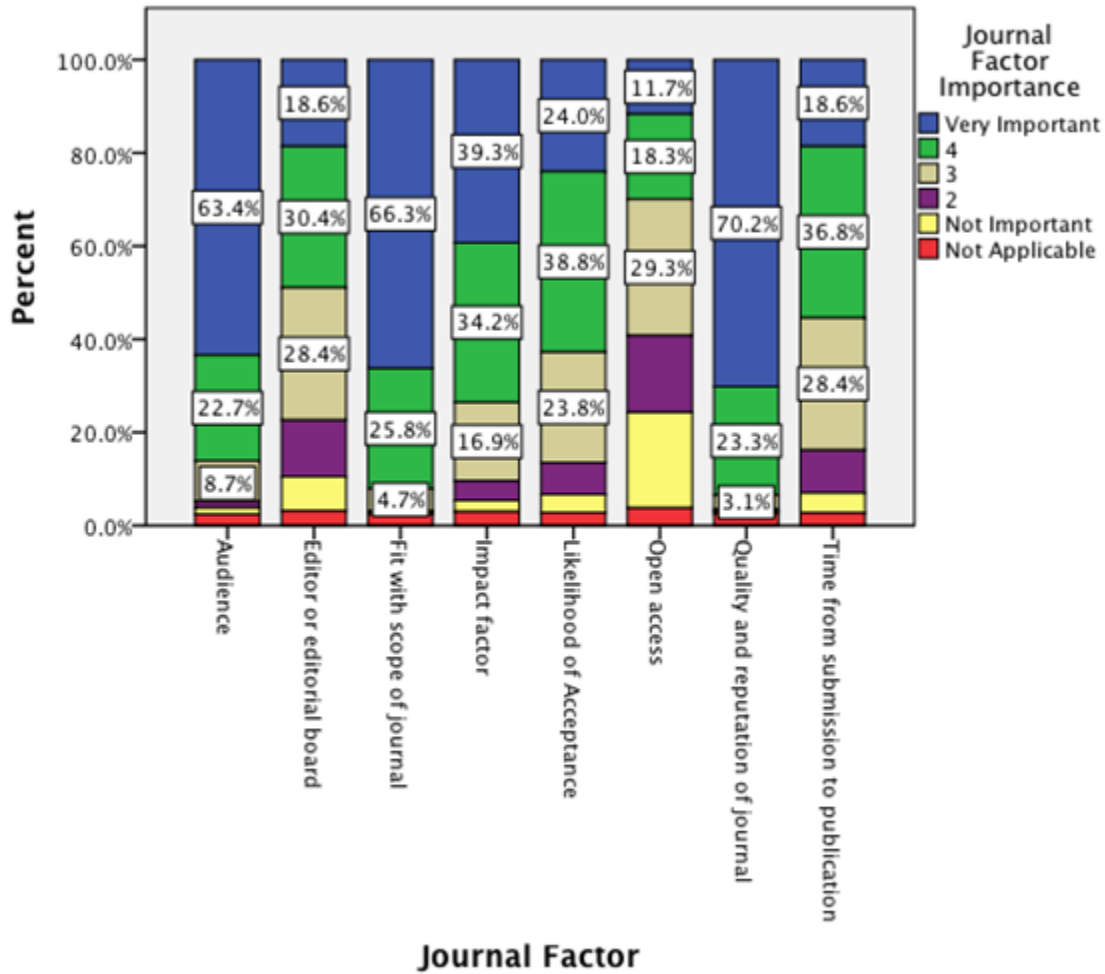


Figure 1: Author Factors in Selecting a Journal

It is perhaps not surprising that graduate student researchers had the same top journal-related priorities as faculty and postdoctoral researchers. However, they tended to rate the factors of quality/reputation, fit, and audience lower in importance relative to faculty and postdoctoral researchers.⁸ Additionally, a number of graduate student respondents expressed in open-ended comments that the choice of where to submit was not theirs. Similarly, respondents from life sciences/medicine, physical sciences, social sciences, and engineering/computer sciences all stated that their advisors' or PIs' recommendations were an additional consideration in

⁸ Tenopir, C., Dalton, E.D., Fish, A., Christian, L., Jones, M.K., & Smith, M. (2016). What motivates authors of scholarly articles? The importance of journal attributes and potential audience on publication choice. *Publications*. Under peer review.

choosing an outlet. Some may have rated these top journal factors lower because the choice of publication is not theirs and they may be interested in simply getting something published.

There was a striking lack of variation across subject disciplines in the high level of importance given to reputation, fit, and audience. This consensus points to the universal importance of being part of a high-quality conversation among peers. When the survey team examined mid-range factors, including impact factor, likelihood of acceptance, time from submission to publication, and editor/editorial board, these are still considered somewhat important across all position types and subject disciplines. Compared to the other positions, faculty rated impact factor and likelihood of acceptance as less important. Faculty, particularly in certain fields, may feel that impact factor does not necessarily equate to quality or visibility within a specific field, and it is not necessarily an indicator of fit for an author's work. As more experienced researchers, they may also feel more confident about the likelihood of getting their work accepted by the right outlets. Graduate students, on the other hand, are often inexperienced and may just want their work to be accepted somewhere.

There was also significant variation across subject disciplines in terms of the importance of these mid-range journal factors. Those in life sciences/medicine, for instance, placed higher importance on impact factor, likelihood of acceptance, and time from submission to publication than other subject disciplines. Taken together, these three factors may indicate a more calculated, quantitative way of accounting for research productivity in the life sciences field. It could also represent an interest in rapidly circulating scholarly findings. Those from mathematics also have distinct thoughts on the importance of journal factors, although due to their small sub-sample size ($n=45$), the findings are often non-significant. For instance, mathematicians rated likelihood of acceptance and editor/editorial board higher than anyone else, while caring the least among subject disciplines about impact factor and time from submission to publication. This finding may be reflective of a close-knit field with relatively few journal outlets in which to publish.

Finally, open access was rated the lowest in importance across all position types and subject disciplines. Within both independent variables, however, there are differences in the level of importance assigned to this factor. Faculty members rated open access as less important than either graduate students or postdoctoral researchers. It is possible that faculty are both more informed and more opinionated about the open access model (or at least more comfortable in expressing their opinions); the strength of their opinions shows in that a full 26.5% of faculty rated the importance of open access as "not important". Graduate students' more moderate rating of the importance of open access may stem from uncertainty about the model.

However, this difference may also indicate a generational shift in attitudes toward the open access model. Graduate students from across multiple disciplines (including arts/humanities, engineering/computer science, life sciences/medicine, and social sciences) commented that

everyone should have access to all research. This finding may indicate an emerging attitude in a new generation of researchers who have stronger beliefs about open access than their predecessors, or it may merely be expressing a utopian rather than a pragmatic opinion at an early stage in their career.

In terms of subject discipline, differences in the perceived importance of open access may be attributable to familiarity with or availability of quality open access journals in a given field. For instance, the lowest rating comes from those in social sciences (including business, education, and law), where open access journals may not carry the same prestige as traditional gated-access journals. Indeed, previous research has found that, at least from among those in the field of business, open access journals are seen as lacking prestige, and there is a perception that publishing in them would be damaging to a scholar's career.⁹ Those from engineering/computer science give open access the second lowest rating from among subject disciplines, which is consistent with previous research that recorded similar hesitation.¹⁰ In addition to the availability of these types of journals, the relatively low importance of open access may also be a reflection of the perceived value of wide accessibility to published research.

The survey also explored issues pertaining to the funding of gold OA publishing. In general, respondents were reluctant to pay author charges to publish their papers. This finding is especially true for those from the arts/humanities, who may have less funding from which to draw. Researchers from traditionally better funded disciplines are more willing to pay, and to pay more, for APCs.¹¹ For example, those from life sciences and medicine were the only respondents willing to pay more than \$2000 per APC.

Overall, the maximum range that respondents deemed to be a reasonable APC amount varied according to the source of funds. For the amount respondents would be willing pay from their personal research funds, the majority of respondents chose "none" (55.2%), followed by "less than \$100" (31.6%). When discretionary research funds was presented as a potential APC source, most selected "less than \$100" (29.4%) or \$100-\$499 (30.6%). The idea of funding APCs with an open access fund through the library was somewhat polarizing, with 31.5% choosing "none" and 25.7% jumping to \$2000-\$2999 as the maximum reasonable APC range for funding from this source. \$100-\$499 was the highest chosen range for both departmental or other institutional funds (30.1%) and grant funds (28.6%).

⁹ Hahn, S. E., & Wyatt, A. (2014). Business faculty's attitudes: Open access, disciplinary repositories, and institutional repositories. *Journal of Business & Finance Librarianship*, 19(2), 93-113.

¹⁰ Mischo, W. H., & Schlembach, M. C. (2011). Open access issues and engineering faculty attitudes and practices. *Journal of Library Administration*, 51(5/6), 432-454.

¹¹ Tenopir, C., Dalton, E.D., Christian, L., Jones, M.K., McCabe, M., Smith, M., & Fish, A. (2016). Imagining a gold open access future: Attitudes, behaviors, and funding scenarios among authors of academic scholarship. *College & Research Libraries*. Submitted.

Author Focus Group and Survey Ramifications for the Model

The survey and focus groups were undertaken to establish a baseline understanding of stakeholders' perception of "value" and potential "buy-in" among scholars with respect to APC-driven open access publishing. Both instruments explored the extent to which the research community would accept changes to the scholarly communication model, and whether this audience could identify and embrace possible benefits of such a shift.

From a consumption standpoint, participants in the focus groups and survey exhibited high satisfaction rates with their level of access through their institutions. Nonetheless, they are aware that access barriers still exist and they recognize that not all academics are as well-situated as they are. From the reader's point of view, open access was roundly praised, with the benefits of access regardless of institutional membership being highly touted. This indicates that the benefits of open access for the overall scholarly ecosystem may be useful tools to highlight the value of a publishing model shift.

From the authoring perspective, opinions on open access varied greatly. Much concern was focused not only on the financial implications of the APC model on the participants themselves, but also on departments, disciplines, positions, universities, and countries with less money than themselves. Concerns regarding the quality of open access were also a consideration, sometimes with the charge of APCs being attached to the stigma of "pay to play." Attitudes in medicine and the health sciences regarding APCs were more open, with those areas being more accustomed to paying those fees in order to publish. Also in these subject areas, an expectation of openness has been established via NIH and other funders' access requirements. Social scientists and humanists did not, on the whole, have as much experience with publishing in open access journals, partly because there are not as many open access journals in their particular fields. This lack of exposure, combined with the absence of a page charge culture in these disciplines, are likely contributing factors to the higher degree of suspicion toward open access that was observed among the social sciences and humanities. Open access skepticism in these fields was exacerbated by perceived limitations in personal and external grant funding that could be used to defray APC costs.

Although many participants across disciplines were supportive of the open access model, there was concern about identifying funding models that do not put direct pressure on the author, are easy to implement, have an application process for funds that would be fair to all and without barrier to smaller institutions, and seek to address the huge range in APCs from journal to journal. A clear, equitable, and easy-to-implement solution was obviously desirable. It is reasonable to conclude that the success with which such models are identified will influence the degree of buy-in open access will find within the authoring community.

Participants also demonstrated a preference for the library to play an active coordinating role in any transition to an open access model. In particular, social scientists and humanists believed that the library should step in for researchers, especially for those in less well-funded disciplines (such as their own). They do not have as many resources to draw upon as so-called hard sciences. They apparently trusted the library in the role more so than the university; as one social scientist said, applying for university-level grants was a waste of time. Those in the fine arts and humanities worried that drawing from university funds may adversely affect any other funding for their departments. Participants from medicine and engineering supported the library's role in gold open access as well. Some from medicine wanted APCs to be part of the library's negotiation with publishers, while those in engineering believed that libraries should pay APCs instead of subscriptions.

Although no participants from the humanities or fine arts mentioned the use of external grant funding to cover APCs, all other disciplines mentioned this resource as a viable alternative to the library or the university. A few social scientists said that APCs should be part of hiring packages (employment benefits). Other social scientists mentioned treating gold OA like a commercial marketplace in which publishers have to compete for authors' business. These strands are consistent with this project's modelling, which looks at how to draw from various funding sources and provide authors with some form of "buying power" to create a functioning gold open access market.

b. Publisher Survey

Rationale

To better understand publisher attitudes and activities related to open access, the project team designed and executed a survey of the Association of Learned and Professional Society Publishers' (ALPSP) publisher members. The survey was constructed to establish baseline understandings of (1) How trends in APCs have impacted current operations and planning for publishers; (2) Publishers' degree of confidence in their ability to adapt their operations and business models to a changing environment; and (3) How publishers feel that cultural, jurisdictional, and disciplinary factors are affecting the transition. In particular, we hoped to assess how publishers think their role might change in the event of a scholarly communication shift to an APC-funded model, and what perceived opportunities and challenges such a change might produce.

Methodology

ALPSP was selected as a project partner because it is the largest international trade association of scholarly and professional publishers. ALPSP membership consists of 240 journal publishers from across disciplinary, geographic, and business model spectra. The survey instrument was

developed by the project team, coordinated by Greg Tananbaum. ALPSP provided input into the survey design via their executive leadership team, led by Audrey McCulloch, Chief Executive, and Suzanne Kavanagh Director of Marketing and Membership Services. A copy of the survey questionnaire is included as Appendix C.

SurveyMonkey was used to conduct and manage the survey. Invitations were sent to 240 ALPSP members on May 8, 2015. Two reminders were sent before the survey was closed at the end of the month. Recipients were informed that their responses would be anonymized before sharing with any partner outside of ALPSP, in order to underscore their ability to reply candidly to potentially sensitive strategic questions. Only the ALPSP team and Tananbaum, who is under a separate non-disclosure agreement with ALPSP, have access to the raw results.

Demographics

In total, 71 ALPSP members responded to the survey, a rate of 30%. This was well in excess of our anticipated response rate of 10%. The vast majority of respondents represented not-for-profit publishers.

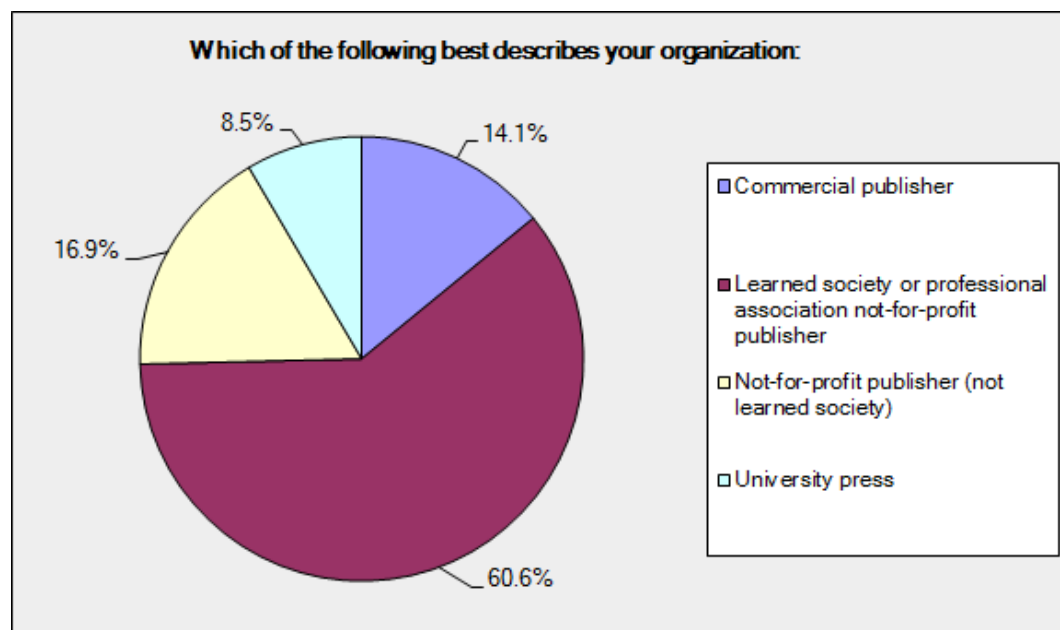


Figure 2: Publisher Survey Demographics

Additionally, responding publishers tended to be modest in size, averaging of 26 full-time equivalent (FTE) employees, 16 directly owned journals, and 9 journals published on behalf of partners. Respondents published across a wide range of disciplines, the most common being biology, humanities, medicine/health, social sciences, and chemistry.

Publisher Survey Findings

Current Business Models and Revenue Streams

Subscriptions are far and away the primary business model (80%). Only 3% of respondents cited OA funded by APCs as their primary model. An additional 8% indicated a combination of subscriptions and APCs. Beyond academic institutional subscriptions, a number of other revenue streams generate a significant amount of income (defined here as >10% of overall journal revenue) for publishers. These include non-academic institutional subscriptions (mentioned by 20% of survey respondents), memberships (18%) individual subscriptions (17%), and advertising (13%).

With respect to open access, 83% of respondents offer an APC model in at least one of their journals, including hybrid options. On average, respondents estimated that 69% of the subscription journals they publish offer a hybrid option, and an estimated 17% of their journals are fully open access. Overall, the respondents estimated on average that 9% of the papers they published in 2014 were open access and paid for by APCs.

When setting APCs, publishers typically do not vary the price depending on the number of pages (92%), or the number of authors (97%). Nor do they typically layer page, color, or other charges on top of the APC (73%). The average APC list price is \$1,980 (all responses were converted to USD), with the lowest price provided as \$0 and the highest as \$5,000. 25% charge APCs for content types other than peer-reviewed papers (e.g., communications, letters to the editor, review articles). Within this group, 62% do not vary the APC price by content type.

The APC pricing strategy is guided almost evenly between “being competitive with the price points of other journals in the field” (36%) and “the calculated cost of publishing an article given your current cost structures, including indirect costs and profit” (32%), with an additional 18% using a combination of both methods. APC pricing is evaluated annually by 43% of respondents, with 38% indicating the pricing is not reviewed according to any formal schedule. Whenever this evaluation is done, the two biggest considerations are the projected volume of papers and their associated projected cost (74%) and staying competitive with the market (71%).

Offering APC waivers to authors who would otherwise have financial difficulties is common practice (68%). Most respondents (83%) indicated that the approximate percentage of authors granted waivers is between 1% and 10%. Nearly the same percentage (79%) indicated that waiver take-up has little or no impact on their overall APC pricing strategy.

Nearly three in four (73%) of responding publishers do not currently provide “offset” arrangements offering credits, discounts, rebates, etc., to universities that pay both subscription charges for publications and fees to make papers open access. Among those that

do not offer offsets, 70% are unsure if they plan to do so in the next one to three years. The remaining 30% are split evenly between “yes” and “no”.

Open Access Trends and Attitudes

23% of responding publishers have converted at least 1 journal from a subscription to a fully open access model. Looking forward, 19% of respondents anticipate they will convert at least one journal from subscriptions to fully OA in the next 1-3 years, while 53% do not. In general, publishers have a balanced view of OA and APCs:

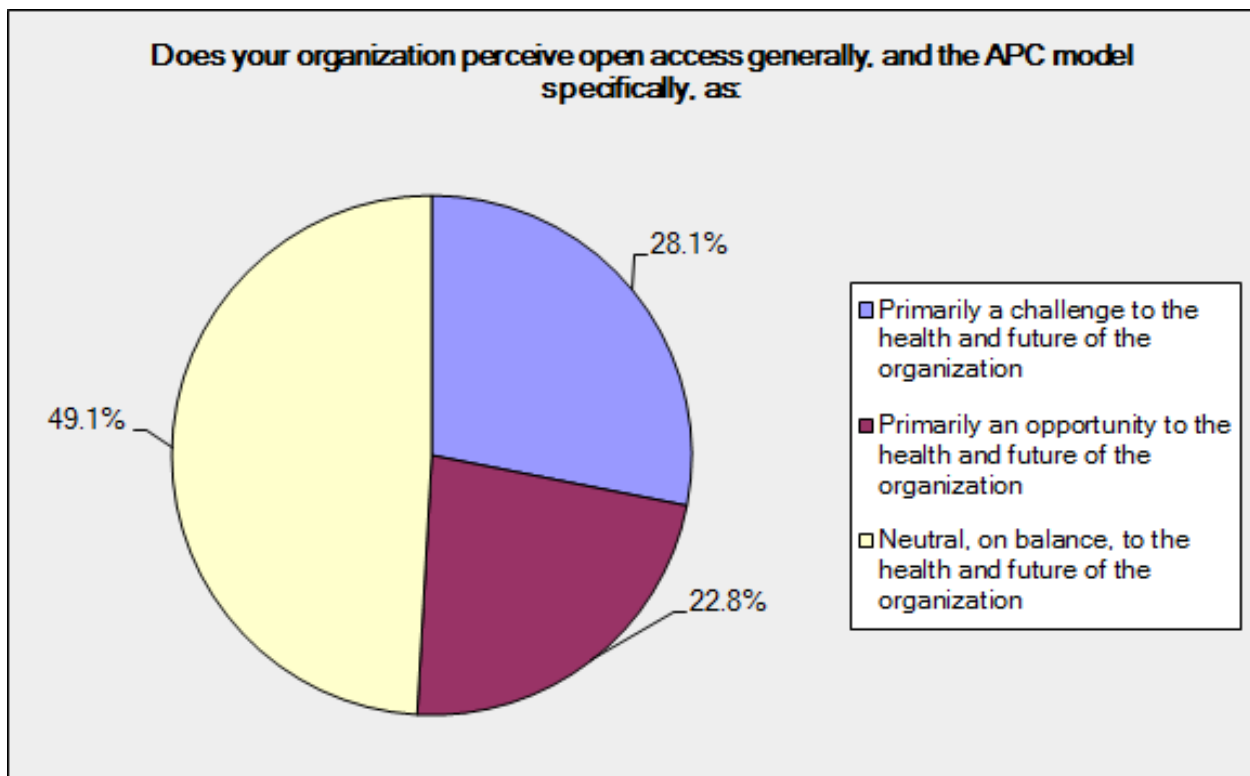


Figure 3: Publisher Views on Open Access

Digging further into these attitudes, the most frequently cited opportunity (14%) was the ability to explore new revenue streams. On the flip side, 25% of respondents indicated that it is potentially challenging to develop financially sustainable OA models. The most common comment about the current rate of OA uptake was that it has been slow and/or modest (63%).

Nearly half of respondents (48%) indicated that the emergence of open access and APCs has had little impact on how they engage with libraries. Close to one-quarter (24%) said that it has expanded what they talk about with institutions, and/or who is involved in that conversation.

The rise of open access has impacted publisher perspectives and business practices in the past three to five years primarily by spurring a re-assessment of business models and practices (45%). OA has also led to the development of new journals and/or new policies for existing journals (30%).

Looking forward over the next three to five years, open access is perceived as most likely to impact organizational infrastructure such as systems and workflows (19%). A similar number of respondents (16%) don't believe OA will cause a significant organizational reorientation. That said, open access seems to be impacting publishers most in their long-term strategic planning, and impacting them least when it comes to concrete financial and personnel allocations:

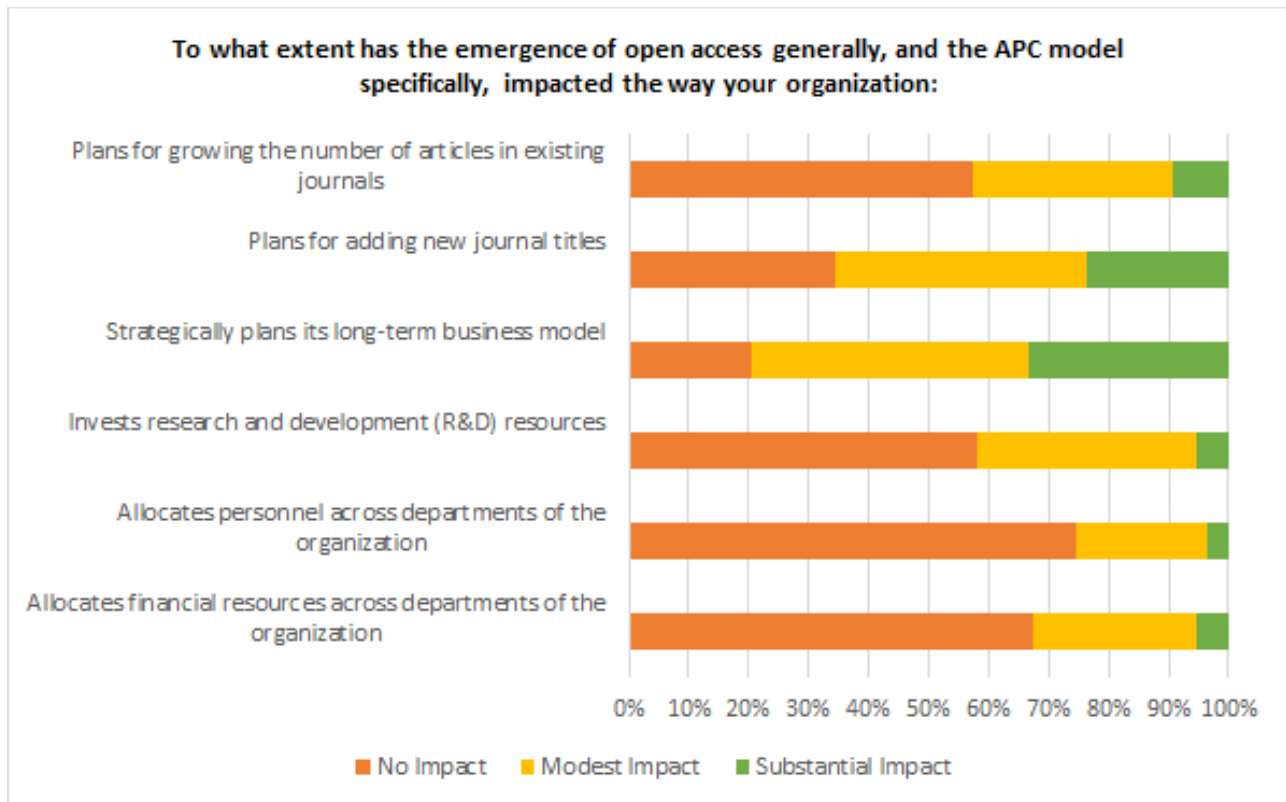


Figure 4: Impact of Open Access on Publisher Activities

Ramifications for the Model

The ALPSP member survey results provided useful insights into publisher attitudes on a number of issues impacting this project's findings:

- *How trends in APCs have impacted current operations and planning for publishers.* As per Figure 4 above, open access developments have had modest or no impact on how publishers allocate financial resources and personnel across departments, how they invest in R&D, and how they are planning to grow their publication outputs among existing journals. The impact of open access on publisher outputs, bottom lines, relationships with libraries, and other markers has not been transformative to date. This may be driven by the relatively modest open access output among survey respondents, and the correspondingly small percentage of revenue derived from open access papers and journals. Open access developments are having a more substantial impact on how publishers are strategically planning their long-term business model and how they are planning to grow both the number of papers and new journal titles they publish. Additionally, publishers are at least somewhat aware that growth in open access may necessitate an adjustment to operations and workflows in the next several years. Overall, the survey results indicate that open access is largely an issue that publishers are tracking over the long-range horizon, with modest impact to immediate-term operations. This suggests that the findings of this project will be met by publishers with a general understanding that successful long-term business planning requires a clear-eyed assessment of gold open access models that are viable for the range of research stakeholders (e.g., publishers, authors, readers, institutions, libraries).
- *Publishers' degree of confidence in their ability to adapt their operations and business models to a changing environment.* With 72% of respondents not viewing open access as a challenge to the health and future of their organizations, it is reasonable to conclude that publishers are relatively unworried by near-term trends in OA. This finding implies both that they have a solid degree of confidence in their ability to adjust to changes brought upon by open access, and that they view these changes as modest today and in the immediate future. Combined with the findings discussed immediately above, this indicates that publishers are not overly fearful of changes open access may bring. It suggests at least a certain degree of receptivity to scenarios that may substantively alter current operating procedures.
- *How publishers feel that cultural, jurisdictional, and disciplinary factors are affecting the transition.* Because the vast majority of respondents are not-for-profit publishers of one form or another, it is difficult to tease apart responses by organization type (e.g., commercial vs. noncommercial). This limits, to some degree, the extent to which cultural differences can be assessed. It is worth noting that very little distinction in

responses can be found based on a publisher's location. This may be due to the international nature of the industry, in which authors, readers, libraries, and other stakeholders cut across national borders. From a disciplinary perspective, publishers whose lists are more than 25% focused in the humanities tend to have less experience with open access. Only four in nine have a fully OA journal, and the percentage of their journals offering hybrid options is slightly lower than the overall average (67% compared to 83%). As with other disciplines, humanities-oriented publishers indicate that open access is impacting them most in their long-term strategic planning, and impacting them least when it comes to concrete financial and personnel allocations. That said, a slightly higher percentage of humanities-oriented publishers (33%, compared to 28% overall) view open access as a challenge, with a similar difference in the percentage viewing OA as an opportunity (11% humanities vs. 23% overall). Publishers whose lists are more than 25% focused in the social sciences are similar in many ways to the humanities responses, though this is driven in part by the overlap between the two sub-groups. However, none of the nine social science-oriented publishers view open access primarily as an opportunity to the health and future of the organization, while 56% viewed OA as a challenge. These data points indicate that any scenarios requiring a substantial reorientation of the publishing model may need to address a discipline-specific lack of experience with open access, along with the uncertainty and apprehension that may accompany this lack of experience.

V. Quantitative Data Components of the Model

a. Publishing Output Data

Rationale

Given the core task of the project--to model an APC-centric publishing environment--it was crucial to obtain as accurate and complete an understanding as possible of the publishing outputs associated with each partner institution that might be subject to an Article Processing Charge. An important data component of the project involved gathering and analyzing the volume of relevant scholarly publications produced by authors at each partner institution, as well as key characteristics of these publications (such as discipline, author affiliation and role, open access status, and grant sponsorship). These data would allow the modeling team to determine, under various scenarios and within different disciplines, how many APC payments each partner institution would be responsible for on behalf of its authors.

Methodology

The publication data used to construct our models was obtained via partnerships with Thomson Reuters and Elsevier, producers respectively of the two major bibliometric databases Web of Science and Scopus. Together, they index the bulk of the scholarly journal literature produced today. Scopus indexes a larger number of scholarly journals than Web of Science (as of late 2014 when our project was in planning, some 20,000 titles compared to a little over 12,000 titles in Web of Science), including a greater number of journals in the humanities and social sciences. Web of Science, by contrast, contains more comprehensive sponsored research information, particularly in the sciences; Web of Science funder acknowledgement data goes back to 2009 in the sciences, whereas these data were not recorded in Scopus at all during the period under study. Both sources index conference proceedings in addition to journals. Combining these two sources allowed us to assemble a comprehensive picture of the research output of our university partners, including the key characteristics needed for analysis. These sources also formed the basis for gathering library expenditure data from our university partners (see “Library Expenditure Data” section).

Both Elsevier and Thomson Reuters constructed a series of reports for use in our analysis in response to specifications built by the project team (see Appendix F). Data were gathered for the years 2009 through 2013 to allow for five full years of longitudinal bibliometric data. Because there is a lag between when papers are published and when they appear in these databases, and because the project commenced in early 2015, data from 2014 were not included. Thomson Reuters shared with us as well a raw, article-level dataset of all publications in Web of Science’s Science, Social Science, and Arts and Humanities editions. The granularity of this dataset allowed for in-depth analysis of our partner publications. From these several

sets of data, we developed a profile of the relevant publishing output of each institution over the five-year period in a range of broadly-defined subject categories.

Scope of the Data

The bibliometric data available through Scopus and Web of Science were used to define the scholarly publications considered in-scope for the project. Both Elsevier and Thomson Reuters provided title lists from their databases to allow us to determine the set of journals and other publications to consider, as well a list of document types used to classify documents at the article-level. Based on this information, we elected to include:

- Articles (including review articles) in journals other than trade journals covered by either database in any of the years from 2009 through 2013.
- Conference papers either in journals, book series, or dedicated conference volumes covered by either database in any of the years from 2009 through 2013. Because conference papers are a primary form of research communication in certain disciplines, notably computer science and engineering, and because they are often (but not exclusively) published in journals and can be subject to an APC, we considered conference publications integral to our study.

In consultation with our contacts at Thomson-Reuters and Elsevier, some publications were eliminated from our analysis despite data being available:

- Articles and other documents appearing in trade journals: trade journals typically do not publish scholarly research, which is the focus of this study.
- 657 journals covered by Scopus for which Scopus obtained its data solely from Medline. Medline-sourced citations in Scopus often contain incomplete metadata, most commonly missing author affiliations, which were central to our analysis. Out of the 5,000 titles for which Scopus obtains citation data from Medline, approximately 4,000 are also covered by direct agreements with publishers, through which Scopus obtains more complete metadata; another 77 titles had complete metadata as part of the Web of Science dataset. An analysis of the remaining 657 journals suggested that they are far less likely to contain articles from authors at our partner institutions than other Medline journals that are fully indexed in Scopus and Web of Science. To avoid confounding our analysis with questionable data, we chose to exclude these 657 titles from our analysis of publication output and library expenditures.
- Documents in Web of Science where the document type is either “Article;Book”, “Article;Book Chapter”, or “Review;Book Chapter”, and documents in Scopus where the

source is classified as a “Book-in-series” or “Book” and the article is classified as anything other than a conference paper. Books and book series were only included in our data requests to capture as many conference papers as possible. Other forms of book publication, such as monographs, were considered out of scope for this project.

In total, approximately 23,000 journals, 650 book series, and 31,000 conference titles were considered to be in scope for the project, comprising approximately 7.9M worldwide documents in the Web of Science dataset and 10.7M worldwide documents in the Scopus dataset.

Data Gathering

The data-gathering component of this process involved identifying and defining all of the potentially relevant data points and classifications to be used for analysis. The modelling team consulted with technical staff from each bibliometric partner to determine what data were available and how content and data descriptors had evolved in these databases over time. Based on these discussions, the following data elements were used as project inputs.

Journal-Level Data

Basic bibliographic data, including source type, source title, title abbreviation, ISSN, and publisher were available from both providers; also included in the data from each provider was an indication of whether a given journal is open access, based on its inclusion in the Directory of Open Access Journals (DOAJ). These data were largely consistent between the two providers, and were used to merge the two source lists into one, for synthesis of these two separate datasets and in requesting library expenditure data for in-scope materials (see “Library Expenditure Data” section). In addition, both providers included their proprietary subject coding, assigning one or more subject to each journal; see “Defining Disciplinary Classifications” later in this section for a description of our methodology in reconciling classifications between the two providers.

Article-Level Data

Thomson Reuters supplied the modeling team with an article-level dataset from which we conducted most of our analyses. This dataset provided standard bibliographic data for each article, including article type, title, author list, volume, issue, and year published, as well as data which corresponded with the journal-level data, including journal title, publisher, ISSN, and journal abbreviation. The dataset also included a field for “Reprint Author” that identified the author designated as the corresponding author for the paper, as well as the address and affiliation provided by that author for contact purposes.

Several levels of subject classification were included for each article, including Web of Science's proprietary "category" and "research area" classifications, assigned at the article level, as well as a separate subject classification used in Thomson Reuters' Essential Science Indicators (ESI) product. The ESI category is assigned at the journal level for most articles, but is assigned at the article level when the journal is categorized as Multidisciplinary. Additionally, articles were flagged to indicate whether they are indexed in the Arts and Humanities Citation Index (AHCI) as a further point of disciplinary identification. A discussion of these subject classifications appears below under "Defining Disciplinary Classifications."

Articles were identified as open access in the dataset based on mapping from DOAJ applied in late April, 2015. The DOAJ mappings take into account the date range of OA content as represented by DOAJ. For example, if DOAJ indicates that a journal started its OA coverage in the middle of our study period, the earlier articles in that journal will not be marked as open access, while the later ones will. Open access articles in hybrid journals for which APCs have been paid are not identified as such in the dataset.

Separate datasets were provided giving a deeper view into the authors and grants on each paper. Because multiple authors could be linked to the same paper, Thomson Reuters provided a separate data file listing each author on each paper, including the author's name, the paper's accession number, and the author's position in the list (first author, second author, etc.) on that paper. Additionally, because multiple affiliations could be linked to the same author, Thomson Reuters provided a separate file of affiliation data, including the institution and address of the author, the paper's accession number, and the author's position in the list (the latter two allowing for the affiliation file to be connected with the author file). We were able to use this data to construct a smaller raw dataset of only articles with an author from a partner institution. Thomson Reuters also provided a separate data file containing detailed grant information where indexed by Web of Science, including the paper number, granting agency, and grant number where available. The presence of a grant acknowledgment statement in a paper was important to our study because many granting agencies will pay publishing costs for the articles resulting from the research they fund. Further details about grant funding data are below.

Aggregated Article-Level Data

In addition to the raw data provided by Thomson Reuters, aggregated reports were obtained from both data providers for use in the project. The reports obtained from Thomson Reuters in general duplicate information found in the raw Web of Science dataset, but these reports offer points of comparison with those obtained from Elsevier, as a raw Scopus dataset was unavailable.

One report provided aggregated data for each partner institution, journal/source title, year, and document type. This data included the total number of papers with an author from the institution, and the number of papers with the corresponding author from the institution. The report also included authorship patterns, including the number of articles with only one author, the median and average number of coauthors per document, the number of articles with only one affiliation, and the median and average number of affiliations per document. All of these data points were provided by Elsevier and replicable in the Web of Science raw data.

Additionally, reports with aggregated article data were generated to help enrich the information we had about journals; these reports included data points such as total number of publications per journal in a given year, both overall and by institution. Along similar lines, Elsevier was able to provide data about the number of non-research articles published in each journal (e.g., letters to the editor, editorials, etc.) which would need to be subsidized by the APCs for the research article content; it could be expected that journals with a higher percentage of editorial content may be forced to charge a higher APC to cover the production cost of this content. The vast majority of journals had a low percentage of editorial content, with about 90% of journals containing less than 20% non-research article content.

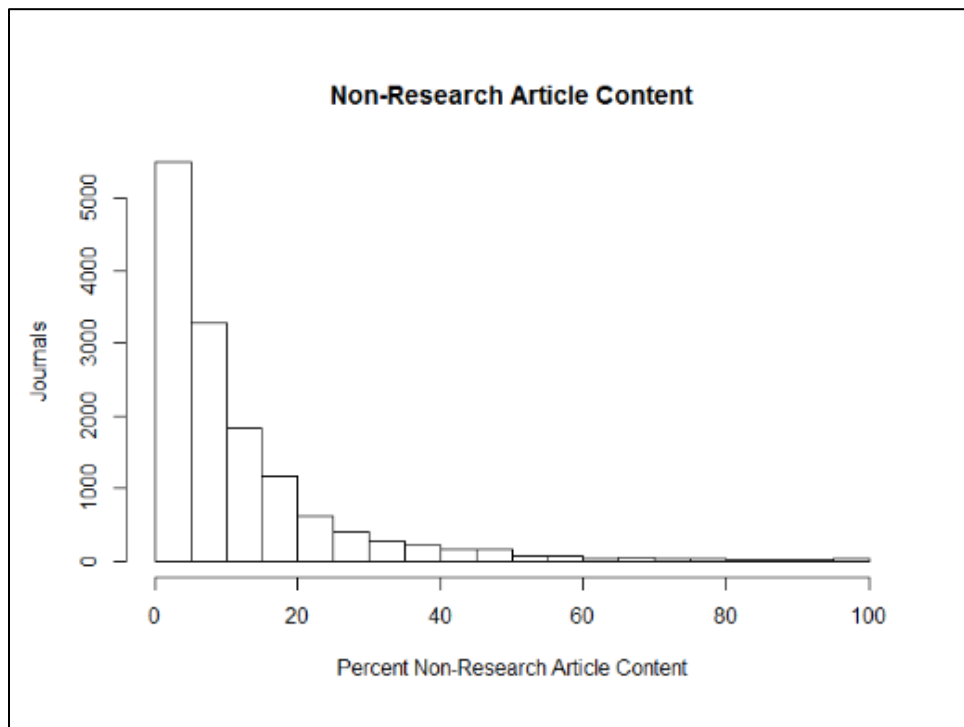


Figure 5: Percentages of Non-Research Article Content within Bibliometric Reports

Defining Disciplinary Classifications

In order to properly conduct a discipline-based analysis of the bibliometric data from Web of Science and Scopus, the modeling team needed to create a high-level subject classification scheme to apply across both data sources. Scopus and Web of Science each use a proprietary classification scheme to assign subjects to journals and articles.

Web of Science has a multi-tiered structure in which each journal (or in some cases, each article) is assigned one or more of approximately 250 categories; each category maps in turn to one of approximately 150 higher-level research areas. Additionally, each journal covered by either Web of Science's Science Citation Index or Social Science Citation Index is uniquely assigned to one of 22 Essential Science Indicators (ESI) subject categories.

Scopus also uses a tiered structure in which each journal is assigned one or more subject codes in Elsevier's All Science Journal Classification (ASJC) system. Each ASJC code maps to one of 27 intermediate-level categories, and each intermediate-level category maps to one of four top-level subjects.¹²

A key goal for a unified scheme was to reconcile differences between the two vendor systems in such a way as to facilitate analysis at an aggregate level while surfacing differences that might be important to the project findings, such as co-authorship patterns or grant funding.

The modeling team created a unified subject classification scheme that correlates the 22 ESI subject categories and the 27 intermediate-level ASJC codes. The resulting scheme, referred to as "PIF Subjects," consists of 13 subject categories. Each ESI and ASJC category was mapped to a single PIF Subject Category, as was the Arts & Humanities Citation Index (AHCI) flag. The resulting PIF Subject mapping appears in Appendix D, and the statistical characteristics observed for each ESI and ASJC category are in Appendix E.

Where an obvious one-to-one correspondence among ESI, ASJC, and PIF categories was lacking, mapping decisions were made based upon statistical characteristics as well as the team members' experience as academic faculty or librarians. For example, the Scopus ASJC category "Chemical Engineering" could have been mapped to either Chemistry or Engineering in the PIF scheme, whereas the ESI category "Neuroscience and Behavior" could be mapped to either "Biomedical Research Disciplines" or "Psychiatry and Psychology." Because the patterns of grant funding and co-authorship for articles classified as "Neuroscience and Behavior" bore more similarity to other biomedical research articles than to articles in psychology and

¹² For a list of Web of Science subject categories, see http://incites.isiknowledge.com/common/help/h_field_category_wos.html. ESI categories are described at <http://ipscience-help.thomsonreuters.com/incitesLive/globalComparisonsGroup/globalComparisons/subjAreaSchemesGroup/essentialScienceIndicators.html>. For a list of Scopus ASJC codes, see the Scopus Source List at <https://www.elsevier.com/solutions/scopus/content>.

psychiatry, we placed them in the PIF “Biomedical Research Disciplines” category. Additionally, many conference proceeding volumes in the Web of Science dataset were not assigned an ESI category; we classified papers from these volumes as “No PIF Category.”

A second challenge in classifying bibliometric data by subject was the presence of multiple subject assignments for articles and journals in both vendor databases. A single journal or article may have as many as four or more subject assignments in either Web of Science or Scopus. In order to develop meaningful subject profiles of article processing charges and other publishing characteristics, it was necessary for our purposes to assign a single primary subject to each journal and article. The Thomson Reuters ESI category, which assigns a single category to each journal (and consequently to each article published in that journal), was therefore given first priority for PIF mapping purposes. Journals and articles indexed in AHCI were assigned to the PIF Arts & Humanities subject even if they also had an ESI category assigned. Journals and conference proceedings that were covered in Scopus but not Web of Science, and which therefore did not have an ESI category or AHCI flag, were assigned a PIF Subject based on ASJC codes. Titles with multiple ASJC codes were assigned the PIF Subject corresponding with a plurality of ASJC codes. If no single PIF Subject was represented by more ASJC codes than the others, a subject prioritization was applied to decide on a code (see Appendix D).

Finally, summary statistics were calculated and assessed to evaluate the mapping and confirm that the ESI categories and ASJC codes being merged in this mapping had similar important characteristics, including authorship patterns and publication growth over time (see Appendix E).

Defining Institutional Affiliations

An important component of the data gathering process involved identifying the set of institutional affiliations within each vendor’s system that best captured the authors from each of our university partners. Each vendor provided us with affiliation data that was reviewed by our library partners for accuracy. Each database treats author affiliations differently and required a different validation process.

For Web of Science, affiliations are based on the address and organization name supplied by the author in the publication metadata. Thomson Reuters connects each address string with a high-level organization (generally the full university), but it does not maintain an authority list of departmental addresses, and many address variants are common. Thomson Reuters generated lists of address text strings associated with each of our partner institutions for review. Scopus, on the other hand, maintains an authority list of affiliations, each mapped to an institution. Affiliation data as it appears in each article is curated and linked to the relevant affiliation authority. This strategy has benefits and drawbacks compared to the Web of Science data. The limited, authoritative list of department, lab, hospital, and institutional affiliations

was easier for our library partners to review; however, without being able to inspect the link between the affiliation text on the paper and the affiliation authority chosen, greater reliance had to be placed on Scopus's affiliation mapping without the ability to apply more granular criteria.

Both Thomson Reuters and Elsevier provided proposed affiliation lists to the project team which we reviewed in collaboration with library contacts at each institution. Through this review, a number of affiliations were removed as being unlikely to be used by researchers at the institution, while others were identified for inclusion. These corrections were returned to the data providers and used to generate the data reports and augment the raw datasets.

While this process was fairly straightforward in many cases, in others (such as affiliated research labs or teaching hospitals) choices had to be made about whether an affiliation included enough institutional authors to merit inclusion in the institution's profile while minimizing the volume of articles published by unaffiliated authors that would then be counted. In general, we chose to err on the side of inclusiveness, with the result that some affiliations have been included portions of whose research might also be funded by other institutions (for example, Dana Farber Cancer Institute was included as a Harvard affiliate, even though not all researchers at DFCI are affiliated with Harvard). This may result in over-counting the research being produced by a given institution, particularly for institutions such as Harvard that collaborate with many local hospitals or other outside organizations. Therefore, the publication volume used for modeling purposes represents more of an upper bound on the institution's output than it does a precise count.

Data Analysis

Assigning Responsibility for Payment

The primary goal of gathering extensive bibliometric data was to measure the publication output of each institution, in order to determine the number of publications for which our institutional partners would have to pay an APC. Cross-institutional collaboration complicates this calculation; the modeling team needed to make a judgment about the most likely methods for assigning payment responsibility in cases of multiple authorship and apply those calculations to our data. Two common methods envisioned for assigning APC payment responsibility are assigning responsibility to the corresponding author, and dividing payment responsibility among co-authors.

The modeling team decided early on that the most likely business model would be to charge the APC to the corresponding author's institution. This reflects current practice for many, if not most, existing open access publishers that charge APCs (examples include BioMed Central, Frontiers, Nature Publishing Group, Wiley Open Access, and Cogent OA) as well as many university open access funds (such as Carnegie Mellon). However, we do recognize that not all

entities currently follow this model; PLOS and Hindawi, for example, charge the submitting author, who may or may not be the corresponding author, while the Harvard Open-Access Publishing Equity (HOPE) fund pro-rates the funding provided based on the number of authors. Dividing payments among all authors could be seen as a fair method, but it has significant logistical challenges, particularly for large collaborations. Assigning payment responsibility to the submitting author would be possible, but is not well tracked in bibliometric databases and so would be difficult to model.

Both Web of Science and Scopus track the corresponding author in their data. Thomson Reuters identifies the corresponding (or reprint) author and that author's cited contact address in its raw data file, while Elsevier was able to include a count of documents with a local corresponding author affiliation for each institution in the reports it provided.

However, some amount of variability exists in the affiliation data. Web of Science quotes a single corresponding author affiliation for each article in its dataset; we may be able to interpret this as the preferred or "home" institution of the corresponding author, and therefore it may be the more likely institution to which the corresponding author charges an APC. However, many of these authors have multiple affiliations in the Web of Science affiliation data file. Thus, two potential methods exist for assigning financial responsibility for a paper to a partner institution through the Web of Science data: assigning it if the partner institution is the "home" institution of the corresponding author (Method 1), or assigning it to a partner institution if the corresponding author has any affiliation with that institution, whether it is the "home" institution or an alternate affiliation (Method 2). This second method results in an increase of between two and six percentage points in the overall corresponding author percentage for a particular partner institution.

Elsevier's method for assigning the corresponding author affiliation in Scopus was less apparent due to the lack of article-level data. However, a spot-check comparison showed that Scopus's corresponding author affiliation most closely matches Web of Science's "home" affiliation, with the additional observation that Scopus is slightly more stringent in its requirement for listing an affiliation as the corresponding author's affiliation. For example, when an author has changed affiliations during the time that a paper is published, Web of Science appears to capture both affiliations and use the newest affiliation as the "home" institution, whereas Elsevier assigns the original listed affiliation as the corresponding affiliation. As a result of this differing approach, Web of Science has a slightly higher corresponding author percentage for each institution under Method 1.

The corresponding author percentages for both Web of Science approaches as well as the Scopus approach can be seen in the table below. This table displays the percentage of papers with an author from each institution where the corresponding author is also from that institution.

Table 4: Corresponding Author Rates, Various Methods

Institution	Web of Science Method 1	Web of Science Method 2	Scopus
Harvard	52%	58%	49%
OSU	58%	61%	56%
UBC	58%	61%	55%
UCB	52%	57%	49%
UCD	56%	60%	53%
UCI	56%	59%	53%
UCLA	54%	58%	51%
UCM	56%	59%	49%
UCR	55%	58%	54%
UCSD	53%	58%	49%
UCSF	49%	54%	46%
UCSB	54%	59%	51%
UCSC	46%	51%	45%

In most cases, throughout the course of our analysis, we have used the higher-percentage assignment method (generally Web of Science Method 2), in an effort to estimate the worst-case scenario for the assignment of payment responsibility to our partner institutions. We acknowledge that the other methods could be closer to the actual assignment of responsibility, but given proper incentives it is feasible that any author with a partner affiliation would go to the partner institution for payment, especially if these institutions are on the forefront of setting policy in an APC-funded business model and therefore offer better resources to their authors.

Grant Acknowledgment Data

An important component of our bibliometric data that required detailed analysis is the acknowledgment of a granting agency in published research. Many funding agencies allow authors to use grant funds to pay for publication charges, and even the Compact for Open-Access Publishing Equity (COPE) calls for granting agencies to pay publication fees when the research being published is grant-funded. While Elsevier was unable to incorporate this information into the reports provided to the project team, Thomson Reuters has been indexing grant acknowledgment statements in Web of Science for the last several years. The Web of Science dataset therefore included good-quality grant acknowledgment data at a paper level,

including both the presence of a grant acknowledgment as well as the granting agency being acknowledged, for disciplines in science, engineering, and medicine.

For disciplines in the social sciences and humanities, neither Thomson Reuters nor Elsevier was able to provide reliable grant acknowledgment data for the time period of the study. However, Thomson Reuters did begin consistently indexing grant acknowledgment statements in Web of Science for articles in the social sciences in mid-2014. The project team obtained a supplementary dataset of all articles indexed from January to October 2015, including the journal in which the article appeared (from which we could determine the PIF Subject of the article) and whether the article had a grant acknowledgment. We then used these data to calculate estimated grant-funding percentages by discipline and applied these percentages to our existing dataset. A table of the percentage of articles that acknowledge a grant, by discipline, in each dataset, is below. The table displays the percent of articles in each subject with a grant acknowledgment statement indexed by Web of Science in our 2009-13 partner institution dataset and our supplementary 2015 worldwide dataset.

Table 5: Grant Acknowledgment Rates by Subject

PIF Subject	Grant Acknowledged - 2009-13 Partner Institutions	Grant Acknowledged - 2015 all
Arts and Humanities	6%	10%
Biomedical Research Disciplines	85%	79%
Business and Economics	8%	35%
Chemistry	88%	83%
Clinical Medicine	62%	52%
Earth Sciences	86%	81%
Engineering	66%	71%
Life Sciences	82%	77%
Mathematics	69%	69%
Multidisciplinary	89%	89%
No PIF Category	1%	0%
Physics and Astronomy	86%	81%
Psychiatry/Psychology	33%	50%
Social Science	32%	42%

As expected, the 2015 grant acknowledgment percentages diverge significantly from our 2009-2013 dataset in the four PIF Subjects within the realm of social sciences and humanities

(highlighted in red text). For three of these subjects, Business and Economics, Psychiatry/Psychology, and Social Science, we chose to use the percentage from the 2015 dataset in our modeling, to correct for the fact that Web of Science did not fully index grant acknowledgments before 2014. Therefore, all of our analyses assume that 35% of articles in Business and Economics, 50% of articles in Psychiatry/Psychology, and 42% of articles in Social Science reflect grant-funded research. This approach is justified by the fact that our remaining partner data, with a very few exceptions, either matches or exceeds the grant acknowledgment percentages in the global dataset.

The Arts and Humanities PIF Subject was treated differently, because grant acknowledgment statements are not reliably indexed in the Web of Science Arts and Humanities Citation Index. The few articles classified as Arts and Humanities that contain grant acknowledgment statements are also indexed in the Social Science Citation Index or the Science Citation Index, and may reflect multidisciplinary work. Additionally, we can see from our research funding and expenditure data (see “Research Funding Expenditure Data” section) that research expenditures in these fields are dwarfed by other disciplines: in the United States in FY2013, research expenditures in these fields constituted less than 1% of all research expenditures. Taking these factors together, we determined that the best approach would be to assume that no article in the Arts and Humanities PIF Subject would have grant funding available to pay the APC. This assumption is reflected in all of our analyses.

Almost all science, engineering, and medicine fields actually show a slightly lower grant-acknowledgment rate in the 2015 dataset. In general, we trust that the indexing Web of Science performed on grant acknowledgement statements is accurate, and we use the article-level data for all of these fields. We posit that the reason for this difference is that our partner institutions are research-intensive North American institutions and their authors are therefore more likely to obtain and acknowledge grants than the general population of worldwide authors represented in the 2015 supplementary dataset. The notable exception is Engineering, which has a higher grant-acknowledgment rate in the 2015 dataset. However, this is due to the composition of the Engineering articles: in the 2009-2013 dataset, 10% of Engineering articles were conference papers, while in the 2015 dataset, less than 1% of Engineering articles were conference papers, perhaps because there is a longer delay in indexing these types of documents. Conference proceedings have a lower percentage of articles with grant acknowledgment statements, thereby explaining this discrepancy.

Synthesizing the Datasets

Web of Science and Scopus offer different levels of coverage and data points; as described above, Scopus offers wider coverage of journals, while Thomson Reuters was able to provide article-level data from Web of Science that included data on grant acknowledgments and

offered greater flexibility in determining the affiliation of the corresponding author. In order to build a robust dataset on which to perform our analyses, the data from the two sources had to be synthesized to create a single view of the publication output of the partner institutions. Because Web of Science offered article-level data including several data points that were unavailable through Scopus, we used the Web of Science article-level data as the basis of the synthesized dataset, and incorporated data from Scopus where needed as a supplement.

The first step in this process was to identify the overlap in coverage between the two databases. For journals, this was relatively straightforward: journals were matched by ISSN; unmatched journals were compared and additional matches were performed by title (usually with brief checks on other fields, such as publisher, to ensure that the match was appropriate). Article counts in each journal were compared to ensure that both databases were providing comparable data. Journals covered by Scopus but not Web of Science were considered to represent unique data points and were used to extend the dataset.

For conference proceedings, the process was more difficult, as there is no standard unique identifier for conferences. Additionally, conference titles have many potential name variants, so matching on exact conference names yielded a much smaller overlap than actually exists. Instead, a process was run to identify key abbreviations cited in conference names, which generally represent the sponsoring or publishing organization (such as IEEE, MRS, AIAA, etc.). The high frequency of a particular organization name was interpreted to indicate that the source covers that organization's conference proceedings. Web of Science and Scopus had several high-frequency organization names in common, so these proceedings were assumed to be part of the overlap in coverage; all others were considered to represent unique items.

The journals and conference proceedings in Scopus that do not overlap with Web of Science comprise approximately 9% of the total articles in Scopus that are attributed to authors at our partner institutions. We refer to this unique Scopus subset from this point forward as the Scopus-not-Web of Science, or SNW, dataset. The data include publication counts by authoring institution, year, document type, and journal/conference proceedings title.

For these articles, we made estimates about grant acknowledgment and corresponding author rates based on calculations from the Web of Science data, to allow us to merge these data with the existing Web of Science data. We applied grant acknowledgement percentages derived from Web of Science to each journal or conference proceeding in the SNW dataset to estimate the number of articles that acknowledged a grant. It is useful to think of this process in the sense of probability: if a particular mathematics journal published three Ohio State University-authored articles in 2013, we would assume, based on the table in the previous section, that there is a 69% chance that each article has a grant acknowledgment statement. This calculates to an expected value of 2.07 (3×0.69) papers with a grant acknowledgement statement; this is the number we use in our modeling.

The rate of papers with a corresponding author from the partner institution was applied similarly, but in a more granular way, to determine how many papers from the SNW dataset we expect to have a corresponding author from one of our partner institutions. As was discussed previously, Scopus records corresponding author affiliations slightly differently from the Web of Science dataset, and we wanted to replicate our Web of Science methodology for the SNW data. The corresponding author percentage in Web of Science is therefore calculated for each discipline, document type, year, and institution, and applied to the SNW dataset accordingly. Because some of these percentages may be calculated on a very small number of data points, we set a minimum threshold in this calculation below which we use the Scopus corresponding author data rather than the Web of Science percentage. For our analyses, we chose a threshold of 20 observations. Additionally, we decided against applying these rates to conference papers and instead used the Scopus corresponding author affiliations for these documents. While we were unable to compare conference proceedings documents directly in a systematic way, the overall corresponding author rates for conference papers in Web of Science were significantly higher than those for Scopus (much higher than for other document types), and so applying those rates to the SNW dataset would risk misrepresenting the Scopus data.

Using these corrections, the SNW data can be analyzed along with the Web of Science data to create a more complete view of the output of each partner institution. The table below shows the volume of publications in each PIF Subject from the Web of Science data and the SNW data.

Table 6: Bibliometric Data Set: Document Counts, by PIF Subject

	Web of Science				Scopus-Not-Web of Science				Increase over Web of Science
	All Docs	Docs w/Grant	Corresponding Authors	Corr Authors w/Grant	All Docs	Docs w/Grant	Corresponding Authors	Corr Authors w/Grant	
Arts and Humanities	6400	0	5605	0	928	0	796	0	15%
Biomedical Research Disciplines	52601	44912	33008	27991	1309	1118	779	665	2%
Business and Economics	5628	1987	3481	1229	964	340	563	199	17%
Chemistry	14887	13089	10426	9233	573	504	435	383	4%
Clinical Medicine	81635	50311	50448	29453	10348	6377	5979	3685	13%
Earth Sciences	17475	15072	9755	8421	1236	1066	680	586	7%

	Web of Science				Scopus-Not-Web of Science				
Engineering	23425	15554	15320	9995	13253	8800	8346	5542	57%
Life Sciences	33616	27756	21407	17611	526	434	317	262	2%
Mathematics	5494	3809	3446	2250	501	347	296	205	9%
Multidisciplinary	1785	1588	1160	1008	256	228	131	116	14%
No PIF Category	14662	163	13825	88	1015	11	264	3	7%
Physics and Astronomy	32197	27713	17103	14576	2278	1961	1275	1097	7%
Psychiatry/ Psychology	13959	6952	8145	4056	245	122	136	68	2%
Social Science	20065	8367	13096	5461	3786	1579	2414	1007	19%
Total	323,829	217,272	206,225	131,372	37,218	22,887	22,411	13,818	11%

Size of the Web of Science dataset and the added data from Scopus, by PIF Subject.

b. Library Expenditure Data

Rationale

The core goal of this project is to evaluate the sustainability of a publishing model based on APC payments rather than subscriptions. Since subscriptions are the primary mechanism for funding scholarly publication, the redirection of funds from institutional subscriptions toward OA support was a foundational assumption in our project. It was important to determine exactly what our partner institutions spent on subscriptions within the scope of the project, to understand how much funding could be redirected to cover, or subsidize, APCs for authors at each institution. Additionally, to allow for more granular analysis, we wanted to identify various components of library subscription expenditures that might change in different ways in an APC-funded system. These components include subscriptions to print materials, subscriptions negotiated by consortia on behalf of the library, and memberships or subsidizing payments to open access publishers. Finally, overall budget allocations were needed to estimate the growth over time of library budgets in an effort to assess whether the typical rate of growth in library budgets would be able to accommodate the rate of increase in research publication.

Methodology

Data were obtained directly from contacts at each partner institution's library in response to data-gathering instructions composed by the project team. Institutions were provided with lists of in-scope materials, and were asked to provide their overall spending on these materials in each year from 2009 through 2013 (see Appendix G). Institutions were not asked to provide itemized costs for each subscription or subscription package, as this could potentially violate non-disclosure agreements between libraries and publishers. Instead, libraries were asked to total their subscription costs locally and only provide the aggregate expenditures to the Pay-It-Forward project team. Assistance with matching against the title list and with identifying out-of-scope materials was provided to library partners on request. Although the project plan originally envisioned developing a data collection template for each library partner to use, a standard process was abandoned in favor of customized assistance because libraries used various integrated library systems and stored their data differently in those systems.

Scope of the Data

The scope of the subscription data gathered for the project was defined by the journal and conference titles for which bibliometric data was available through the project's data partners, Elsevier and Thomson Reuters. An in-scope title list was generated by merging coverage lists from Scopus and Web of Science for the years included in the study; separate lists were generated for journal titles, book series titles, and conference proceedings. Additionally, because many journals are purchased in packages rather than individually, a list of potential content packages was generated using the SFX KnowledgeBase, the link resolution database in use at the University of California, which contains content lists for hundreds of publishers and information providers. Publishers and packages which included in-scope materials were identified in the KnowledgeBase by matching against a list of ISSNs from the merged title lists. The ten University of California campuses were asked to exclude any subscriptions paid for and managed by the California Digital Library since data on these expenditures were already available to the project team.

Subscriptions and other expenditures were considered to be out of scope if they were not the publisher's own platform or the primary source from which libraries would hold an ongoing subscription to published content. Resources such as abstracting and indexing services, aggregators such as JSTOR and LexisNexis, and one-time purchases for journal backfile content were all considered to be out of scope for the project. The list of publishers and publisher packages was vetted to remove any of these out-of-scope resources, in order to assist our library contacts with generating an accurate list of in-scope spending.

As with the bibliometric data gathered from Web of Science and Scopus, data were gathered which covered subscriptions from calendar year 2009 through calendar year 2013. Since the

vast majority of subscriptions for a particular calendar year are paid within several months on either side of January 1, we chose to gather data by fiscal year. Therefore, financial data for the fiscal year running from July 2008 to June 2009 was taken to represent subscriptions for calendar year 2009, and so on. All U.S. campuses followed the July-to-June fiscal year; the University of British Columbia followed an April-to-March fiscal year, but library data contacts there were able to extract spending on a July-to-June schedule, to make their data comparable with the rest of our partners.

Data Gathering

Lists of in-scope titles and publishers were distributed to libraries, and, as described above, libraries were asked to provide us with the aggregate spending on these resources for each year from 2009 through 2013. For most libraries, this involved generating a list of their local subscriptions, matching that list against the PIF title lists, and totaling the expenditures for each title that matched. Additionally, libraries were asked to provide lists of the titles and publisher packages included in these totals; we then checked that no out-of-scope resources were included in the totals.

In addition to aggregate expenditures for all subscriptions, libraries were asked to divide their totals among certain characteristics, to allow for more flexibility in predicting what funding is available to cover the institution's APC costs. These characteristics included:

- **Format.** Where possible, libraries were asked to divide their yearly totals into expenditures for print materials, electronic resources, and combined print plus electronic subscriptions. In an APC-funded model, libraries would be able to cancel all electronic subscriptions, but print subscriptions may need additional consideration depending on how they evolve in an OA world (this is discussed further below). Having totals for each format would allow us to evaluate the impact of discontinuing or retaining print. Libraries were also requested to provide title lists for each format.
- **Consortial Negotiations.** Subscriptions licensed through consortia have particular attributes that may affect the transition to an APC model. For example, because consortia negotiate subscription fees with publishers based on the aggregate value of their members' subscriptions, they have greater leverage to control costs than an individual library may have on its own, resulting in lower costs over time relative to the general marketplace. The models outlined later in this report do not consider the possibility of consortial APC negotiations, and so do not reflect any comparable discounting. Additionally, some consortia contribute central monies toward negotiated license fees in order to extend access to content to all consortium members; whether these consortia would find it equally justifiable in an OA model to contribute toward the

publication costs of their members as opposed to procuring broader access as they do now, is an open question beyond the scope of this report to assess. Finally, the way consortial license fees are shared across consortium members may in some cases be different from the prices that these libraries would have paid as individual customers, further impacting the comparison of current to future costs. In order to assess these impacts, it was important to see which subscriptions and how much spending were the result of consortial negotiations.

- **Open Access Memberships and Subsidy Payments.** Library open access memberships and subsidy payments (e.g., via designated open access funds) represent additional institutional expenditures currently going into the publishing system, and thus also form part of the total amount of library expenditures available for redirection to an APC-funded business model. We recognize that some open access funds derive from institutional sources outside the library, such as a Provost or Office of Research, and thus are not, strictly speaking, library expenditures. However, given the modest level of such funds to date, we have chosen to treat them as library expenditures for the purposes of this analysis. Additionally, open access memberships - which generally involve a basic “bulk” payment to the publisher that procures a discounted APC for individual authors - may also represent a potential open access business model moving forward, as they have proven sustainable for some publishers which employ them, and so this data could contribute to our analysis of future publishing models.

Finally, libraries were asked to provide their overall collection budget allocations for each year, along with a general narrative describing major changes in their collections and their collection spending over the five years of the study. Allocations represent a more reliable view of year-to-year funding capacity than expenditures, because expenditures can include one-time windfalls or transfers from other budget lines. This was intended to provide a more stable view of funding changes over the period of our study.

Data Analysis

The need for additional analysis on the data received from libraries was minimal, as in general the data came in a form that was readily usable. Analysis of this dataset was limited to ensuring that the data were complete and were as accurate as possible, and to comparisons among and between the datasets to ensure consistency. Some libraries were unable to provide perfect datasets for each category of data requested. For example, Harvard operates in a decentralized library system, and not all expenditures are entered into the central ILS, creating some potential variability in its spending data (all major journal packages were accounted for,

so the variability is likely to be relatively small). UC Berkeley, for another example, was unable to separate spending on print-only materials from spending on print-and-electronic materials.

In order to facilitate comparisons among datasets, financial data from the University of British Columbia were scaled by the average USD/CAD exchange rate over each fiscal year. Exchange rates were calculated using www.oanda.com.

Consortial Spending

For almost all library partners, the majority of spending on resources within the scope of the project (66% in 2009 to 71% in 2013) was negotiated through a consortium. Harvard is a member of the NorthEast Research Libraries consortium (NERL); the University of British Columbia is a member of the Canadian Research Knowledge Network (CRKN); The Ohio State University is a member of OhioLINK; and the University of California campuses acquire most of their large digital resources through the California Digital Library (CDL). While neither NERL nor CRKN contribute financially to their member libraries' subscriptions, OhioLINK and CDL both do contribute some level of funding, primarily for the purpose of providing "cross-access" to other consortium members: CDL contributes an additional 13% on top of the costs allocated to campuses, whereas OhioLINK contributes roughly 15%. For modeling purposes, our project assumes that these expenditures would not be directed toward APCs in a future gold OA environment, because the incentive of a "cross-access" subsidy would no longer be relevant. The larger implications for the role of consortia in an APC model are discussed in the "Final Conclusions and Implications" section.

Print Subscriptions

On the whole, the percentage of expenditures going towards resources in print decreased over the course of the study, especially from 2009 to 2010, when many institutions reduced their library budgets in response to the financial crisis. Despite this decrease, partner libraries, on average, reported that in 2013 approximately 11% of their expenditures on in-scope resources went towards subscriptions for materials in print (4%) or in combined print and electronic (7%) formats. Percentages for each partner in 2009 and 2013 (aggregating all UC campuses into one) are below.

Table 7: Library Expenditures on Print Resources

	2009		2013	
Partner	Percent of expenditures on print subs	Percent of expenditures on p+e subs	Percent of expenditures on print subs	Percent of expenditures on p+e subs
Harvard	7%	9%	4%	7%
OSU	11%	4%	4%	4%
UBC	11%	5%	3%	3%
All UCs	10%	12%	4%	8%

While these percentages are relatively small, they still deserve consideration for their effect on the overall financial structure of the model we are envisioning. Two future scenarios can be envisioned for print: either print will completely disappear, and libraries will rely on digital preservation; or publishers will continue to offer print (largely because of ancillary demand from other sources), and some libraries will choose to acquire it. In our modeling, we assume that the first scenario is more likely for the institutions in our study, and we allocate all current spending on print materials toward APCs. However, we recognize that if print continues to be available and libraries choose to retain some print subscriptions, the money available to the APC-based model would need to be reduced by a modest percentage.

Expenditures by Discipline

A subsidiary goal of the project was to examine the percentage of library journal expenditures committed to resources in different subjects in order to illuminate discipline-specific patterns in an institution's spending and compare these patterns to predicted costs in a fully APC-funded system. This can help institutions understand where they may need to commit additional support (see the "Financial Model" section for this comparison). We recognized in the planning process, however, that publisher packages would make this analysis difficult. These large agreements, particularly multidisciplinary packages such as those from major commercial publishers, required further title list analysis to determine the appropriate percentage of overall expenditure committed to each discipline. Many libraries lacked an accurate link between their orders/expenditures database and their bibliographic database, and so were

unable to provide complete title lists for their package subscriptions, much less a cost breakdown at the journal level.

CDL, however, has been maintaining a database of journal titles in each publisher package for years that includes an internal journal-level cost assignment used to assess journal value. This cost, internal to CDL, is determined by a combination of factors including individual journal expenditures rolled forward from the inception of a license, and proportional assignments for packages that are not based on prior subscription history. The journal-level cost can be interpreted as the fractional value that CDL assigns to each journal in a package. By mapping PIF subject into these title lists, we can see the proportion of expenditures committed to supporting each discipline. The table summarizing this disciplinary split for UC system-wide packages in 2013 is below. Amounts have been converted to 2015 USD.

Table 8: UC System-Wide Content Package Expenditures, 2013

Subject	2013 Spend <i>(thousands of US dollars)</i>	2013 Percentage
Arts and Humanities	\$834	4%
Biomedical Research Disciplines	\$3,318	14%
Business and Economics	\$603	3%
Chemistry	\$2,658	11%
Clinical Medicine	\$2,326	10%
Earth Sciences	\$1,547	7%
Engineering	\$2,766	12%
Life Sciences	\$2,933	13%
Mathematics	\$738	3%
Multidisciplinary	\$104	<1%
Physics and Astronomy	\$2,709	12%
Psychiatry/Psychology	\$770	3%
Social Science	\$1,631	7%
Out of Scope	\$539	2%
Total	\$23,476	100%

While all UC campuses subscribe to a reasonable number of journals individually outside of the CDL-negotiated packages, this disciplinary split accounts for a large majority (78%) of all UC

expenditures on journals in the scope of the project (ranging from a low of 71% to a high of 89%).

While the CDL breakdown offers a defensible approximation of disciplinary spend across journal packages, we should also note that allocating journal level costs to large ‘big deal’ packages is more art than science. Some publishers license journal content as a database, making any per-title cost allocation difficult if not impossible; in other cases, publishers and customers may disagree about the cost basis for individual journals within long-standing packages. Our investigation confirmed our original sense that analyzing journal expenditures at a disciplinary level was fraught with complications and could easily consume a separate, dedicated project. For this reason, our discussion of disciplinary distinctions in the “Financial Model” section of this report should be taken as indicative rather than definitive.

Total Funds for Redirection

In all, the primary goal of the library expenditure data collection process was to determine the amount of funding available from the library which could be redirected from subscription and other payments toward paying APCs on behalf of affiliated authors. After evaluating the data that were collected, the funds we determined to be available for redirection included all subscription costs for in-scope materials (including those for print subscriptions) and all OA publisher memberships and subsidies. Additionally, we added to this total any funding which had directly contributed to APC payments via library-administered open access funds; these data were collected as a part of the APC data collection process and is discussed in the “APC Data” section. The total funds available for redirection by each partner library for the year 2013 are in the table below. Amounts have been converted to 2015 USD, based on CPI data gathered from <http://www.usinflationcalculator.com/>.

Table 9: Total Redirectable Library Expenditures, 2013

Institution	Journal/Package Expenditures <i>(thousands of US dollars)</i>	OA Memberships <i>(thousands of US dollars)</i>	APC Payments <i>(thousands of US dollars)</i>	Total Funds for Redirection <i>(thousands of US dollars)</i>
Harvard	\$8,198	\$0	\$16	\$8,214
Ohio State	\$5,524	\$5	\$0	\$5,529
University of British Columbia	\$6,192	\$48	\$0	\$6,240

Institution	Journal/Package Expenditures <i>(thousands of US dollars)</i>	OA Memberships <i>(thousands of US dollars)</i>	APC Payments <i>(thousands of US dollars)</i>	Total Funds for Redirection <i>(thousands of US dollars)</i>
UC Berkeley	\$5,573	\$20	\$86	\$5,679
UC Davis	\$3,990	\$11	\$18	\$4,020
UC Irvine	\$3,413	\$10	\$41	\$3,464
UC Los Angeles	\$4,143	\$14	\$0	\$4,157
UC Merced	\$629	\$5	\$8	\$642
UC Riverside	\$2,207	\$6	\$0	\$2,213
UC San Diego	\$4,200	\$16	\$0	\$4,216
UC San Francisco	\$1,497	\$14	\$20	\$1,530
UC Santa Barbara	\$2,791	\$5	\$16	\$2,812
UC Santa Cruz	\$1,559	\$5	\$0	\$1,565
Total:	\$49,916	\$159	\$206	\$50,281

Collection Budget Growth

The collection of five years' of data allows a longitudinal view of trends in overall collection budget allocations and in funds available for redirection. A summary of these data, aggregated for all partners and normalized for inflation to 2015 USD, is below.

Table 10: Combined Partner Library Budgets and Redirectable Expenditures

Year	Collections Budget Allocation		Funds Available for Redirection	
	<i>Allocation</i> <i>(thousands of US dollars)</i>	<i>Percent Change</i>	<i>Available Funds</i> <i>(thousands of US dollars)</i>	<i>Percent Change</i>
2009	\$151,973	N/A	\$52,792	N/A
2010	\$153,042	0.7%	\$50,538	-4.3%

Year	Collections Budget Allocation		Funds Available for Redirection	
	<i>Allocation (thousands of US dollars)</i>	<i>Percent Change</i>	<i>Available Funds (thousands of US dollars)</i>	<i>Percent Change</i>
2011	\$155,256	1.4%	\$51,146	1.2%
2012	\$151,330	-2.5%	\$50,666	-0.9%
2013	\$151,217	-0.1%	\$50,281	-0.8%

While this longitudinal data does not depict an obvious trend, it is worth noting that the 2008 financial crisis immediately preceded our study period. The notable reduction in funds available for redirection from 2009 to 2010 presumably represents the budget cutbacks required of many of our partner libraries at that time. Following a small bounce-back from 2010 to 2011, both the funds available for redirection and overall allocations remained essentially flat, with an overall slightly downward trend.

c. Research Funding Expenditure Data

Rationale

Another assumption of this project, based on analysis performed in the planning phase, was that grant funds would play an important role in helping to underwrite publication costs. While the use of grant funds to subvent publication is often the subject of some controversy, much APC-funded publication relies heavily on grant funding today.¹³ As we explore below, many federal granting agencies in North America classify publication charges as an allowable expense, and this is increasingly true of private funders as well. While historically these expenses have been in the form of ancillary costs such as page and color charges, such funds are increasingly being applied to fund open access charges.

While the grant acknowledgment statements in our bibliometric dataset (see “Publishing Output Data” section above) were used to estimate the percentage of articles for which grant funding was available and thus the total amount of grant funding that might be allocated for this purpose, it was important to place this within the context of the overall grant funding available to the institution. The data described here allow us to assess the impact that allocating grant funds to APCs would have on overall institutional grant budgets and are thus a key component in assessing the viability and sustainability of any model that utilizes external

¹³ Solomon, D. J. and Björk, B.-C. (2012), Publication fees in open access publishing: Sources of funding and factors influencing choice of journal. *J. Am. Soc. Inf. Sci.*, 63: 98–107. doi: 10.1002/asi.21660

grants as a source of funding. These relationships are explored later in the “Model Viability” section.

Data Gathering

All research funding and expenditure data were compiled from publicly available sources. Data were gathered for fiscal year 2013 and compared to the bibliometric and library expenditure data from that year. Research expenditures for U.S. institutions (by institution, discipline, and funding source) were culled from the Higher Education Research and Development Survey (HERD), an annual survey conducted by the National Center for Science and Engineering Statistics that is the primary source of information on R&D expenditures at U.S. colleges and universities.¹⁴ Expenditure data are preferable to award data for the purposes of this analysis, in part because they are closer in time to both the actual work being performed and the publication of any related research.

Canada collects data on institutional research expenditures comparable to the HERD survey, but its reports are aggregated at the province level rather than by institution. Therefore, we gathered research funding--not expenditure--data for the University of British Columbia for FY2013 directly from the UBC’s Research Funding and Statistics page (<https://research.ubc.ca/research-excellence/research-funding-statistics>) .

Additionally, to investigate the degree to which grant funding can be used to support APC payments, and with assistance from staff at the University of British Columbia, we investigated the funder policies for the largest granting agencies in the U.S. and Canada, including the National Institutes of Health, National Science Foundation, Department of Energy, Department of Defense, and NASA in the U.S., and the Natural Sciences and Engineering Research Council of Canada, Canadian Institutes of Health Research, Michael Smith Foundation for Health Research, and Canada Foundation for Innovation in Canada.

Data Analysis

Research Expenditure Summary data, by Total and Discipline

Our U.S. partner institutions included four of the top ten universities in the HERD survey in total research expenditures for FY2013, with three more in the top 25; a chart of total research expenditures for our partner universities is below.

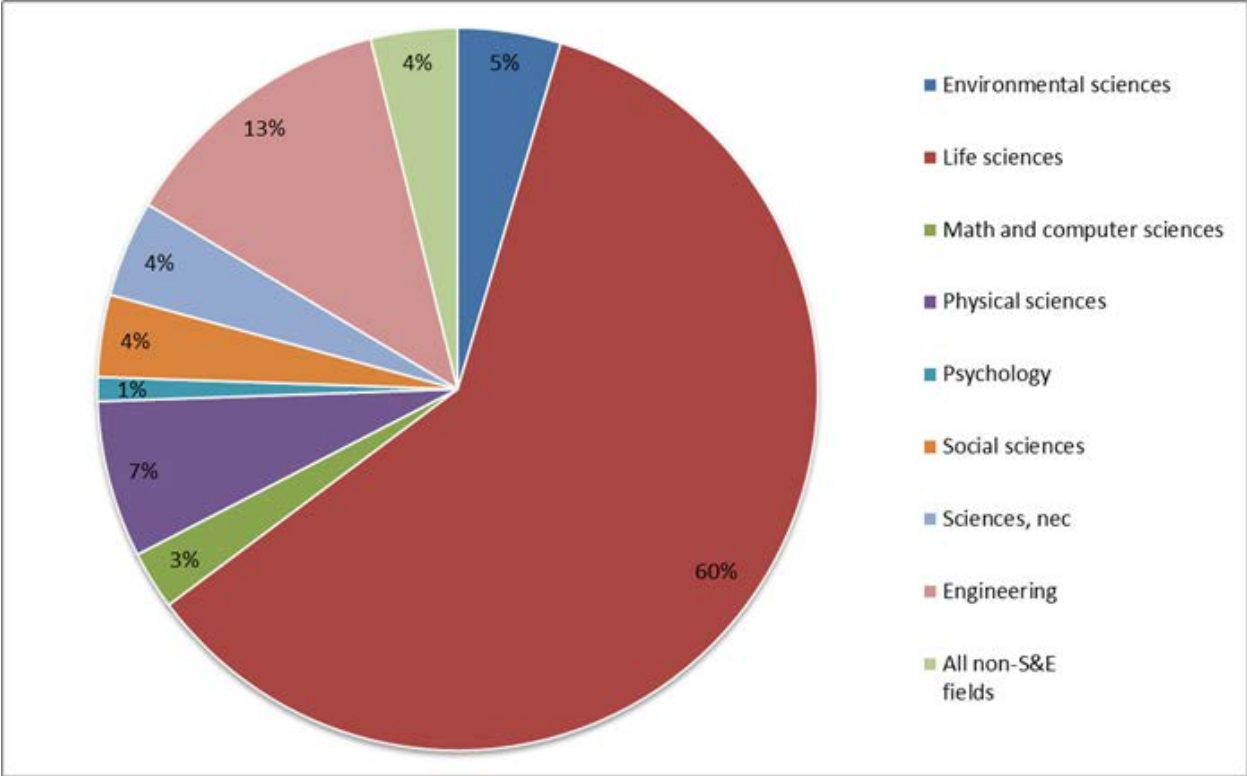
¹⁴ See <http://www.nsf.gov/statistics/srvyherd/>. A senior research analyst in the UC Office of the President assisted the project team in analyzing and interpreting the HERD data.

Table 11: U.S. Partner Research Expenditures

Rank	Institution	Expenditures, in Thousands (2013 USD)
5	U. CA, San Diego	\$1,075,554
6	U. CA, San Francisco	\$1,042,841
7	Harvard U.	\$1,012,766
10	U. CA, Los Angeles	\$966,659
20	OH State U.	\$793,373
24	U. CA, Berkeley	\$727,002
25	U. CA, Davis	\$725,734
65	U. CA, Irvine	\$347,773
87	U. CA, Santa Barbara	\$236,497
121	U. CA, Santa Cruz	\$150,777
131	U. CA, Riverside	\$132,617
253	U. CA, Merced	\$22,909

Source: 2013 HERD survey, table 18

In total, our 12 partner institutions included in the HERD survey (Harvard, Ohio State, and the ten University of California campuses) accounted for over \$7 billion in research expenditures in 2013, or 11% of the U.S. total for that year. Sixty percent of these were in the life sciences:



Source: 2013 HERD survey, table 19

Figure 6: U.S. Partner Research Expenditures by Discipline

These percentages are generally in line with funding patterns across the HERD survey, implying that our partner institutions are, on the whole, representative of the research profile of most U.S. institutions. These discipline-specific percentages are also reasonably well aligned with the publication output of our partner institutions, indicating that no particular subject is receiving significantly more funding per publication than any other. Exceptions to that are in the Physical Sciences and Psychology, where the share of research funding is notably less than the share of publications.

Table 12: Partner Research Funding vs. Publication Output, by Discipline (2013)

PIF Subject	HERD Subject	Papers	Research Funding
Biomedical Research Disciplines Clinical Medicine Life Sciences	Life Sciences	50%	60%
Mathematics Engineering (incl. Computer Science)	Math and Computer Science	12%	15%

PIF Subject	HERD Subject	Papers	Research Funding
	Engineering		
Chemistry Physics and Astronomy	Physical Sciences	14%	7%
Earth Sciences	Environmental Sciences	5%	5%
Psychiatry/Psychology	Psychology	4%	1%
Social Science Arts and Humanities Business & Economics	Social Sciences (incl. Economics) All non-S&E (incl. Business)	10%	8%
Multidisciplinary No PIF Category	Sciences, nec	5%	4%

Note that the HERD subject categories and PIF Subjects do not line up perfectly, as HERD categories were not taken into account in the creation of the PIF Subject scheme. In this table we have endeavored to make the most accurate combinations possible, based on our knowledge of the subject schemes. For example, Computer Science is a part of Engineering in the PIF subject scheme, but the HERD subject combines it with Mathematics. To account for this, we have combined Math-CS-Engineering in this table.

Only a very small percentage of research expenditure - less than 1% - is in the fields of arts and humanities (a subset of the “All non-S&E fields” slice in the above chart); more than half of these are from internal institutional funds (see “Calculating outside funding totals” below). . In the humanities, publications resulting from a grant are also more likely to be monographic in nature rather than in the form of journal articles and conference proceedings, which are the subject of this study. For these reasons, and in the absence of grant acknowledgement data for these fields from our bibliometric sources, we have made the conservative assumption in our modeling that no grant funding is available to support journal publication in the arts and humanities (see the “Publishing Output Data” section). Given the minor amount of journal publication affected -- accounting for only 2% of all publications from our partner institutions -- discounting grant funding in these fields does not have any appreciable impact on our findings.

Calculating Outside Funding Totals

Based on the HERD data that were gathered for U.S. institutions, and incorporating the research award data made available by the University of British Columbia, we generated a summary of extramural funding in order to compare this to the costs that are allocated to granting agencies in our model. In creating this summary table, we removed “Institution Funds” as a source from the HERD data in order to isolate external sources only (UBC did not require this adjustment since institution funds are not reported on its statistics page). Currency

conversions were applied to the data to convert from CAD to USD and from 2013 USD to 2015 USD, to ensure comparability in our implementation examples. The total grant funding summary generated through this process appears in the table below.

Table 13: Research Expenditures from External Sources (2013 expenditures in 2015 USD)

Institution	Research Expenditures (external sources only) <i>(in millions of US dollars)</i>
Harvard	\$762
Ohio State	\$693
University of British Columbia*	\$531
UC System	\$4,508
UC Berkeley	\$583
UC Davis	\$560
UC Irvine	\$296
UC Los Angeles	\$803
UC Merced	\$14
UC Riverside	\$96
UC Santa Barbara	\$203
UC Santa Cruz	\$125
UC San Diego	\$937
UC San Francisco	\$891

*Note: UBC totals are inferred from UBC Research Funding and Statistics page; see document text for detail.

The UBC totals here are not perfectly comparable to the rest of the institutions; as mentioned above, these totals represent awards rather than expenditures, and are somewhat further removed from the actual costs that researchers incurred through their work which were charged back to granting agencies.

It should be noted that the data reported in the HERD survey include indirect cost recovery, that is, the portion of research funding which the institution takes out of each grant award to cover overhead and administration. However, because these totals are being used to determine viability and sustainability of the model for the research enterprise, we have left these indirect cost recovery amounts in the data. As such, these totals represent the actual amounts of money that granting agencies paid to each institution to cover research and related activities, as opposed to the amounts of money that researchers themselves had available to them.

Finally, in removing the institution funds from our datasets to create this summary table, we removed approximately 62% of all funding allocated to arts and humanities, but only 22% of funding allocated to all other disciplines. This discrepancy serves as additional evidence that papers classified with a PIF Subject of Arts and Humanities would have few grant funds available to fund publication.

Funder Policies

Our article-level Web of Science datasets contained not only an indication of whether an article acknowledged a grant, but also identifying data about the grant that was acknowledged. Based on these data, we investigated the funder policies for the most common granting agencies, to confirm whether publication charges are an allowable expense for most grant-funded research.

The National Institutes of Health, National Science Foundation, Department of Energy, Department of Defense, and NASA are the five most common granting agencies for our partner-affiliated publications by far, comprising 72% of the U.S. total. The table below is derived from granting agencies acknowledged in the Web of Science dataset, 2009-13. Note that some papers acknowledge more than one grant; “Any of these 5” is not a direct total of the five agency counts.

Table 14: Grant Acknowledgements by Agency and Institution, 2009-2013

Institution	Papers Acknowledging Grants	Papers Acknowledging:						Percent Acknowledging Any of These 5
		NIH	NSF	DoD	DoE	NASA	Any of These 5	
Harvard	34,237	20,558	3,825	800	695	1,206	24,522	72%
OSU	9,145	3,313	2,230	83	600	301	5,811	64%
UBC	11,576	768	426	36	128	73	1,329	11%
UCB	12,655	2,860	4,346	157	4,199	795	9,666	76%
UCD	10,359	3,555	2,624	97	916	169	6,348	61%
UCI	5,668	2,525	1,841	67	434	247	4,276	75%
UCLA	12,808	6,244	2,900	261	727	689	9,382	73%
UCM	438	91	177	3	44	10	281	64%
UCR	2,904	469	1,305	45	311	60	1,814	62%
UCSB	4,310	642	2,595	45	603	189	3,281	76%
UCSC	2,105	326	1,119	8	287	378	1,607	76%
UCSD	11,687	6,123	2,890	128	707	278	8,798	75%
UCSF	9,576	7,084	365	240	97	18	7,235	76%
Total U.S.	115,963	53,804	26,225	1,934	9,622	4,340	83,043	72%

The obvious outlier in this table is University of British Columbia, which naturally receives few grants from U.S. federal agencies. A separate analysis revealed the top granting agencies for UBC as the Natural Sciences and Engineering Research Council of Canada, Canadian Institutes of Health Research, Michael Smith Foundation for Health Research, and Canada Foundation for Innovation:

Table 15: Grant Acknowledgements by Agency, University of British Columbia, 2009-2013

Agency	Papers Acknowledging	Percent of All UBC Corresponding authored Papers
NSERC	5,471	47%
CIHR	3,606	31%
MSFHR	1,829	16%
CFI	887	8%
NIH	768	7%
NSF	426	4%
Any of these 6*	8,914	77%
Total UBC reprint-authored papers	11,576	

* Note that some papers acknowledge more than one grant; “Any of these 6” is not a direct total of the six agency counts.

All of the top five U.S. funding agencies list publication costs as an allowable expense (as, presumably, do other federal granting agencies governed by U.S. Office of Management and Budget funding policy).¹⁵ Of the four main Canadian agencies, publication costs are an allowable expense for all but the Canada Foundation for Innovation (CFI), which funds infrastructure development rather than research activities.^{16, 17} However, 95% of the 887 articles with a UBC corresponding author that acknowledge a grant from CFI also acknowledge one of the other five major granting agencies in Canada. Based on these data as well as the trend toward support for publication charges on the part of private funders,¹⁸ the general assumption that funder policies allow authors to draw upon grant funds to subsidize publication charges appears to be sound in most instances.

¹⁵ See, in general, [Circular A-21](#) of the Office of Management and Budget, and in particular, [NIH](#), [NSF](#), [NASA](#), [DoE](#), [DoD](#) (see application announcements).

¹⁶ [NSERC and CIHR](#); [Michael Smith Foundation](#) (see p.5).

¹⁷ [CFI Policy and Program Guide](#) (see p.14).

¹⁸ See for example the policies of the [Bill and Melinda Gates Foundation](#).

d. APC Data

Rationale

In order to project an APC-based business model that is both viable and realistic, it was necessary to understand the current environment of APC pricing and payments. These data give a picture of the current volume of open access publishing and the funding allocated to it, particularly in regions (such as certain European countries) where it is becoming more common. In addition, portions of these data support a cost-per-article analysis and provide some insight into the price-setting strategy of open access publishers.

Methodology

Data were obtained from a range of sources, encompassing both APC prices advertised by publishers and actual payment data from libraries, institutions, and other funding organizations which provide financial assistance to authors who want or are required to publish in open access journals. The terms “APC prices” (or “APC pricing”) and “APC payments” are used below to differentiate these two types of data. APC pricing data were gathered for full OA journals only; discussions of hybrid journal pricing are based on earlier studies. APC payment data were collected for both full OA and hybrid OA journals. Data were gathered and, in many cases, synthesized with the bibliometric datasets described in the “Publishing Output Data” section for further analysis.

Data Gathering: APC Pricing Data

Full OA Journals

A previous large scale study of APC prices conducted by Solomon and Björk collected pricing data from journals listed as charging APCs in the Directory of Open Access Journals (DOAJ).¹⁹ A similar study that updated this work was conducted in 2014 by Morrison et al.²⁰ The authors collected data on a stratified (by size) sample of 1,584 journals, or approximately 60% of the journals listed in the DOAJ as charging APCs at the time of the study. Both of these datasets include only full OA journal APCs and not hybrid journal charges. The Morrison study provides the most comprehensive and up-to-date sample of APC charges that we were able to locate. A summary is presented below of the APC charges broken down by the discipline coding used in the study.

¹⁹ Solomon DJ, Björk B-C. (2012) A Study of Open Access Journals Using Article Processing Charges. *Journal of the American Society for Information Science and Technology* 63(8):1485–1495.

²⁰ Morrison H, Salhab J, Calvé-Genest A, Horava, T (2015) Open Access Article Processing Charges: DOAJ Survey Publications 2015, 3(1), 1-16; doi:10.3390/publications3010001.

Table 16: Morrison et al. APC Data by Subject

<i>APC in USD</i>			
Main Subject Classification	Mean	N	Std. Deviation
Agriculture	\$844	40	733.92
Biology and Life Sciences	\$1,588	149	780.03
Education	\$604	17	659.31
General Works	\$786	24	619.42
Humanities	\$909	10	778.10
Law	\$800	1	
Medicine	\$1,462	530	763.39
Physical Sciences and Math	\$808	69	423.40
Science General	\$1,092	10	544.71
Social Sciences	\$781	73	657.29
Technology and Engineering	\$570	123	460.51
Total	\$1,222	1,046	794.27

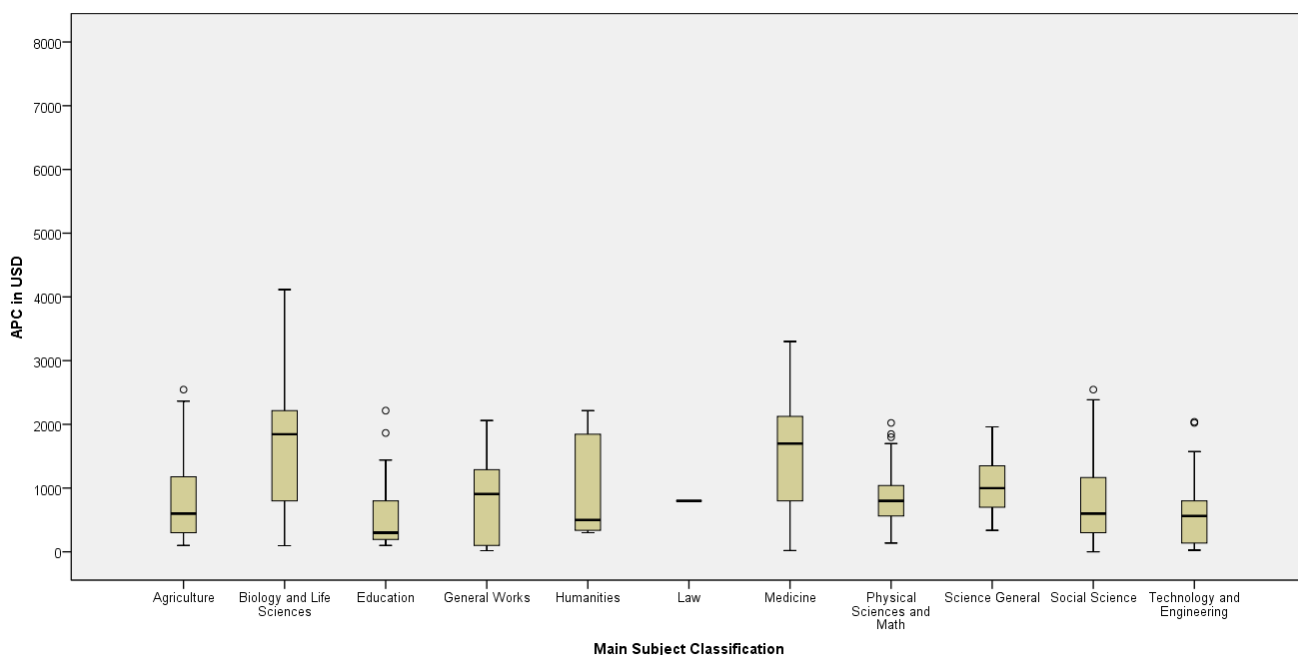


Figure 7: APC Pricing of Full OA Journals, by Subject (Morrison Data)

In early 2016, the team updated the Morrison APC pricing data with more current values for those journals in which authors from partner institutions published most frequently. A set of

316 relevant journals was created by selecting all open access journals in our Web of Science raw dataset that contained at least 10 articles with an author from a partner institution. Updated APCs were obtained for these journals by analyzing the publishers' web pages. Through this work, 282 journals were updated with more current pricing information, including nine journals with alternate pricing methods (generally submission fees rather than publication fees, or publication fees tied directly to the article's page count), 28 journals with no payment required, and 245 journals requiring a standard APC payment. The remaining 34 journals did not mention APCs, a likely but unconfirmed sign that they do not require payment for publication. APCs gathered in this manner were converted to USD using exchange rates from mid-February 2016. These data were used to augment the APC prices obtained via the Morrison study.

Data Gathering: APC Payment Data

Several publication payment databases from European funding agencies have been made public and were harvested for this project. These datasets contain payments and other metadata at the individual publication level, generally including the publications' Digital Object Identifier (DOI). The payments in these databases were in Great British Pounds (GBP) or Euros (EUR). For simplicity, a single currency conversion rate to USD was used of 1.3 for EUR and 1.6 for GBP.

Additionally, many of our partner institutions' libraries have administered open access funds, reimbursing authors who paid APCs to publish in open access journals. We were able to gather a limited amount of data on payments made from these funds for consideration in our study.

The various datasets we obtained are described below.

United Kingdom (UK) Universities

Stuart Lawson and his colleagues at Jisc compiled payment data from a number of UK universities.²¹ We combined two overlapping datasets and removed the duplicates. The payments were converted from GBP to USD using an exchange rate of 1.6 which roughly reflected the exchange rate during the period the APC payments were made. The data include both full OA and hybrid payments.

²¹ Lawson, Stuart (2014): RCUK APC data (2013-14). figshare. <http://dx.doi.org/10.6084/m9.figshare.1180122> Retrieved 18:02, Aug 07, 2015 (GMT).

Wellcome Trust

The Wellcome Trust maintains a special budget for paying publication charges for the research it funds. The Trust has released APC payments made during its 2012-2013 and 2013-2014 fiscal years.²² As with the UK university data above, a currency conversion rate of 1.6 was used for converting from GBP to USD. The data include both full OA and hybrid payments.

German Universities and Foundations

APC payment records were available for 22 German universities and 5 other participating institutions.²³ The payment data in EUR were converted to USD using an exchange rate of 1.3 which roughly reflected the exchange rate during the period the payments were made. Payments were only made for publication in fully OA journals.

Austrian Science Fund (FWF)

The FWF covers the cost of APCs and other publication charges for researchers it funds, and makes payment records available to the public.^{24, 25} The data for 2013 were available at the time we merged the data with Web of Science. Unfortunately 2014 data that became available after we requested Web of Science metadata from Thomson Reuters could not be incorporated. The data include both full OA and hybrid payments.

Partner Institutions

Harvard, The Ohio State University, the University of British Columbia, and five University of California campuses were able to provide data relating to APC payments made by or on behalf of authors. For most partners, this dataset listed actual, article-level payments made by a library-administered fund. These datasets are, in general, consistent in nature with the previously-described European datasets, in that they represent payments made by a funding organization on behalf of authors. Additionally, because they reflect actual payments, they also incorporate institution-specific discounts and other situations in which the payment differs

²² Kiley, Robert (2014): Wellcome Trust APC spend 2012-13: data file. figshare.

<https://dx.doi.org/10.6084/m9.figshare.963054.v1>; Retrieved: 22:24, May 20, 2016 (GMT);

Kiley, Robert (2015): Wellcome Trust open access (OA) spend and compliance monitoring: 2013-14. figshare.

<https://dx.doi.org/10.6084/m9.figshare.1321361.v5>; Retrieved: 22:25, May 20, 2016 (GMT)

²³ *Datasets on fee-based Open Access publishing across German Institutions*. Bielefeld University.

<https://github.com/OpenAPC/openapc-de>; Retrieved: May 10, 2015

²⁴ Reckling, Falk; Kenzian, Margit (2014): Austrian Science Fund (FWF) Publication Cost Data 2013

<http://dx.doi.org/10.6084/m9.figshare.988754>;

²⁵ Reckling, Falk; Rieck, Katharina (2015): Austrian Science Fund (FWF) Publication Cost Data 2014. figshare.

<http://dx.doi.org/10.6084/m9.figshare.1378610>

from the list price APC. Data were requested for 2009, the start of our study period, to the present.

However, these datasets are less conclusive since library payment data reflects a small subset of papers from the institution published in OA journals, and often includes limits on reimbursement. UC Davis, for example, has a \$1000 APC reimbursement limit, and Harvard’s HOPE fund reimburses a prorated percentage of the APC depending on the percentage of authors on the paper with a Harvard affiliation. This situation makes it impossible to assume that the payment accurately reflects the full APC paid for each paper; indeed, with an average significantly lower than what is seen for the European databases, it is likely that these data do not represent the full APC payment for many of the articles listed. The average reported payment, from library-administered open access funds at partner institutions for 2009 to present, is in the table below.

Table 17: APC Payments from Library-Administered Funds at Partner Institutions

Institution	Articles	Average APC
Harvard	60	\$1,015
Ohio State	23	\$870
UC Davis	73	\$945
UC Santa Barbara	32	\$1,524
UC Irvine	30	\$1,395
UC Merced	10	\$1,389
UC San Francisco	27	\$1,257
Total	255	\$1,131

For a few of the reporting institutions, article-level payment data were not available. Instead, the institutions provided estimates of the payments made by authors at their institutions based on bibliometric data and list-price APCs. As a result of these issues, historic library APC payment data ultimately did not inform our model.

Data Analysis

Synthesizing with Web of Science and Scopus Data

To create a more nuanced view of both the APC price data and the APC payment data, the datasets described above were synthesized with the bibliometric data described in the “Publishing Output Data” section in several ways:

1. Mapping PIF Subject to Article-Level Payment Data

The APC payment datasets obtained from the various European databases generally included a DOI for each paper. This allowed us to assign a PIF subject classification to each paper by merging the payment data with article-level data from Web of Science. The merged dataset helped to illuminate discipline-specific differences in current behavior involving APC payments. Since the APC payment data covered 2009 through 2015, whereas our article-level bibliometric dataset only extended through 2013, we provided the payment datasets to Thomson Reuters, whose staff merged the datasets and returned more current information to us. Summaries of payment data for each source are in Table 18. Note that these include only articles which were successfully mapped to a record in Web of Science. Additionally, FWF data from 2014 were not available in time to be included in this mapping.

Table 18: Summary of Compiled APC Payment Data

	Articles: Full OA Journals	Average APC: Full OA Journals	Articles: Hybrid Journals	Average APC: Hybrid Journals
UK Universities	943	\$2,234	2,136	\$2,857
Wellcome Trust	1,050	\$1,997	2,742	\$3,222
German Universities and Foundations	1,676	\$1,665	0	N/A
Austrian Science Fund (FWF)	165	\$1,936	804	\$2,938
Overall	3,834	\$1,907	5,682	\$3,045

Data were divided into APCs for articles in full open access journals, and APCs for articles in hybrid journals. Unsurprisingly, the hybrid averages, which are least affected by normal marketplace forces, hover around the \$3,000 APC charged by many large subscription publishers, with a better than \$1,000 difference between pure OA and hybrid OA overall averages.

Article counts and average APC payments by PIF subject across all European payment datasets are shown in Table 19. In this summary, we can observe some differences in payment among disciplines. For full OA journals, the most expensive average payments are in Chemistry and

Biomedical Research Disciplines; the least expensive average payments are in Business and Economics and Mathematics, with a spread of nearly \$1500 between the highest and lowest averages. Hybrid journal payment averages, by contrast, again exhibit a smaller spread. However, Chemistry has a lower average APC by comparison, while Psychiatry/Psychology has an average APC on par with Clinical Medicine and Biomedical Research Disciplines.

However, many of these disciplines display very few observations, particularly Arts and Humanities, Mathematics, and Business and Economics, while other disciplines have very large samples, such as Biomedical Research Disciplines. This disparity is likely due to differing levels of maturity in OA markets. PLOS One, for example, is one of several well-established OA journals which focus heavily on medicine and life sciences, while fields such as Business and Economics have largely relied on working paper repositories such as SSRN, rather than OA journals, for free distribution of their literature, and so there are very few journals publishing gold OA articles in these disciplines.

Table 19: Average APC Payments across All European Sources

PIF Subject	Full OA Journals		Hybrid Journals	
	Articles	Avg. APC	Articles	Avg. APC
Arts and Humanities	0	N/A	5	\$2,168
Biomedical Research Disciplines	1,450	\$2,036	1,842	\$3,245
Business and economics	4	\$1,416	37	\$2,500
Chemistry	51	\$2,383	420	\$2,861
Clinical Medicine	673	\$1,881	786	\$3,261
Earth Science	173	\$1,580	280	\$2,923
Engineering	104	\$1,685	394	\$2,705
Life Sciences	703	\$1,886	906	\$2,961
Mathematics	5	\$906	54	\$2,610
Multidisciplinary	76	\$1,813	13	\$2,640
Physics and Astronomy	192	\$1,889	246	\$2,567
Psychiatry/Psychology	256	\$1,647	289	\$3,162

PIF Subject	Full OA Journals		Hybrid Journals	
	Articles	Avg. APC	Articles	Avg. APC
Social Science	147	\$1,813	410	\$2,846
Overall	3,834	\$1,907	5,682	\$3,045

Due to these discrepancies in average APCs and the lack of data in certain disciplines, it is difficult to draw any concrete quantitative conclusions about discipline-specific payment behavior from this data. However, these average APC levels, particularly the overall average APC for publication in full OA journals, can serve as a potential baseline cost point at which research institutions in Europe are currently supporting the gold OA model. It is reasonable to expect that these price levels could be indicative of a sustainable average APC for similar institutions in the future.

2. Merging Journal-Level Pricing Data with Bibliometric Data

Merging the APC pricing data with our bibliometric data allows for several additional insights. First, a PIF subject can be assigned to each OA journal, allowing observation of journal price-setting behavior by publishers within different subject areas. Further, we can narrow this pricing data to just the journals that authors from our partner institutions published in, thereby excluding journals of insufficient quality to attract submissions from our partners. The dataset contained 345 non-zero APCs for journals in which our authors published. Because these observations are publisher-set prices that can be presumed sufficient to sustain the publication of a full OA journal, we can draw conclusions about the actual cost of publishing articles. Further discussion of this concept can be found in the “Cost-Per-Article Analysis” section.

Similarly, we can map the APC pricing data to our bibliometric data at the article level, allowing us to see how much money we anticipate was paid for the publication of these articles over the course of our study. As with the journal-level data, we can examine this data by PIF subject to try to identify any discipline-specific patterns. The data for both of these processes is below; note that because Multidisciplinary journals often contain articles with their own PIF subjects, there is not necessarily a direct match between the number of nonzero APC journals and the number of articles in non-zero APC journals (for example, the one Arts and Humanities journal does not necessarily contain 23 articles from partner authors—some of those 23 may also be in Multidisciplinary journals such as *PLOS One*).

While this data mirrors some of the discipline-specific patterns observed in the article-level

payment data, it also exhibits some of the same weaknesses, specifically a lack of observations in the Arts and Humanities, Business and Economics, Mathematics, and Multidisciplinary subjects. Again, the discipline-level distinctions seen here could be used as a general guide, but do not offer any concrete quantitative conclusions for use in our modeling efforts.

Table 20: APC Pricing Mapped to Partner Bibliometric Data

PIF Subject	Nonzero APC Journals	Avg. APC	Articles in Nonzero APC Journals	Avg. APC
Arts and Humanities	1	\$337	23	\$1,517
Biomedical Research Disciplines	77	\$2,223	6,920	\$2,035
Business and Economics	0	\$0	11	\$1,495
Chemistry	8	\$1,216	262	\$1,251
Clinical Medicine	100	\$2,038	4,458	\$1,873
Earth Sciences	20	\$1,162	1,588	\$1,214
Engineering	26	\$1,220	519	\$1,784
Life Sciences	59	\$1,979	3,466	\$2,065
Mathematics	7	\$828	88	\$789
Multidisciplinary	7	\$1,007	525	\$1,482
No PIF Category	0	\$0	15	\$647
Physics and Astronomy	10	\$1,574	999	\$1,898
Psychiatry/Psychology	8	\$1,976	454	\$2,133
Social Science	22	\$1,909	780	\$1,919
Total	345	\$1,864	20,108	\$1,892

Additionally, using this data we can see how many journals our partner institutions' authors published in which either do not charge APCs or operate on a different payment model. The best data for this observation is the set of 316 OA journals in which our partners have published

at least 10 articles; these are the journals for which we have the most information. The count of journals and articles from this set currently published under various models is in the table below.

Table 21: Payment Models for OA Journals Popular Among Partner Authors

Payment model	Journals		Articles (2009-13)	
	Count	Percentage	Count	Percentage
Non-zero APC identified	245	(78%)	19,665	(89%)
\$0 APC identified (free to publish and read)	28	(9%)	893	(4%)
No APC identified (likely free to publish and read)	34	(11%)	1,520	(7%)
Variable payment by page or other unit	6	(2%)	96	(0%)
Payment through submission fees	3	(1%)	35	(0%)
Total	316		22,209	

Most existing full OA journals in which our partner authors publish regularly operate on an APC-funded model. About 20% of the OA journals our authors regularly publish in either state that they do not charge APCs or don't make a statement about APCs at all on their website (generally implying that they do not charge them). These journals account for about 11% of the article publishing volume for the years in our study. As will be discussed further in the "Financial Model" section, these free-to-publish and alternate-model OA journals introduce some level of variability into our results.

3. Mapping Journal Metrics to Journal-Level Pricing Data

In addition to observing journal-level pricing data broken down by discipline, it is also useful to observe how journal-level pricing compares to journal citation metrics such as Thomson Reuters' Impact Factor (IF) and Elsevier's Source Normalized Impact per Paper (SNIP). Both IF and SNIP were available at the journal level from Web of Science and Scopus, respectively, so these metrics were mapped into the journal-level pricing dataset where our authors published. Of the 345 journals in this set, 331 have an assigned IF and 327 have an assigned SNIP. Scatterplots charting the APC against the IF and SNIP are shown below.

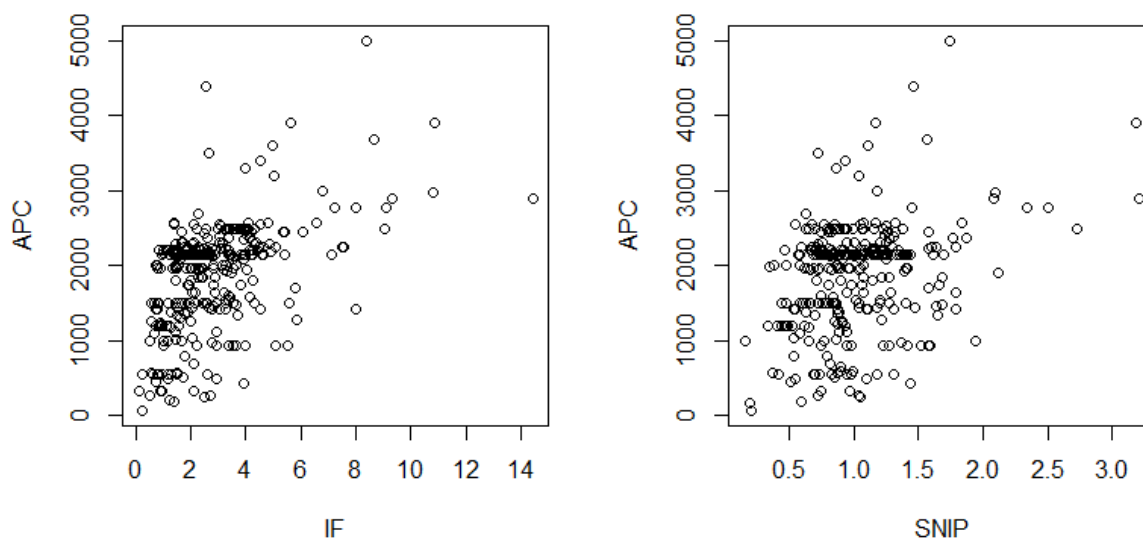


Figure 8: Scatterplot of APC Price vs. Citation Metrics

A Pearson correlation was used to measure the relationships between journal impact metrics and APC pricing observations in the scatterplots below. There was a positive correlation between both citation metrics and price: for IF and APC, the Pearson r correlation coefficient was 0.464 ($p < 0.001$), and for SNIP and APC, the Pearson r correlation coefficient was 0.322 ($p < 0.001$). These values of r both indicate a moderate association between APC pricing and citation metrics. This relationship is explored further in order to predict future APC pricing in the “Financial Model” section.

Additional Analyses

Full OA Journals

As part of an earlier study, Björk and Solomon obtained APC charges for 102 full OA journals published by five large traditionally subscription publishers: Elsevier, Wiley, Nature Publishing Group, Sage and Taylor & Francis. The data were collected in August 2013. These and other large subscription publishers are rapidly expanding their open access portfolios, creating a market that is in significant flux. Between August and December 2013 the number of OA journals published by Elsevier increased from 46 to 72; as of April 2016, Elsevier listed 541 open access journals on its website (see <https://www.elsevier.com/about/open-science/open-access/open-access-journals>). These publishers are both creating new OA journals and, to a lesser extent, “flipping” journals from subscription to OA. Some of these publishers such as Springer (BMC) are acquiring existing OA publishers to expand their OA journal portfolio. The

figure below (reproduced from the Björk and Solomon study) compares the distribution of the APCs from OA-only journal publishers and the full OA journals from traditionally subscription publishers. The average APC levied by the five largest subscription publishers in 2013 was 2,097 USD (i.e. on average 679 USD higher than the APC levied by “born digital” OA publishers).

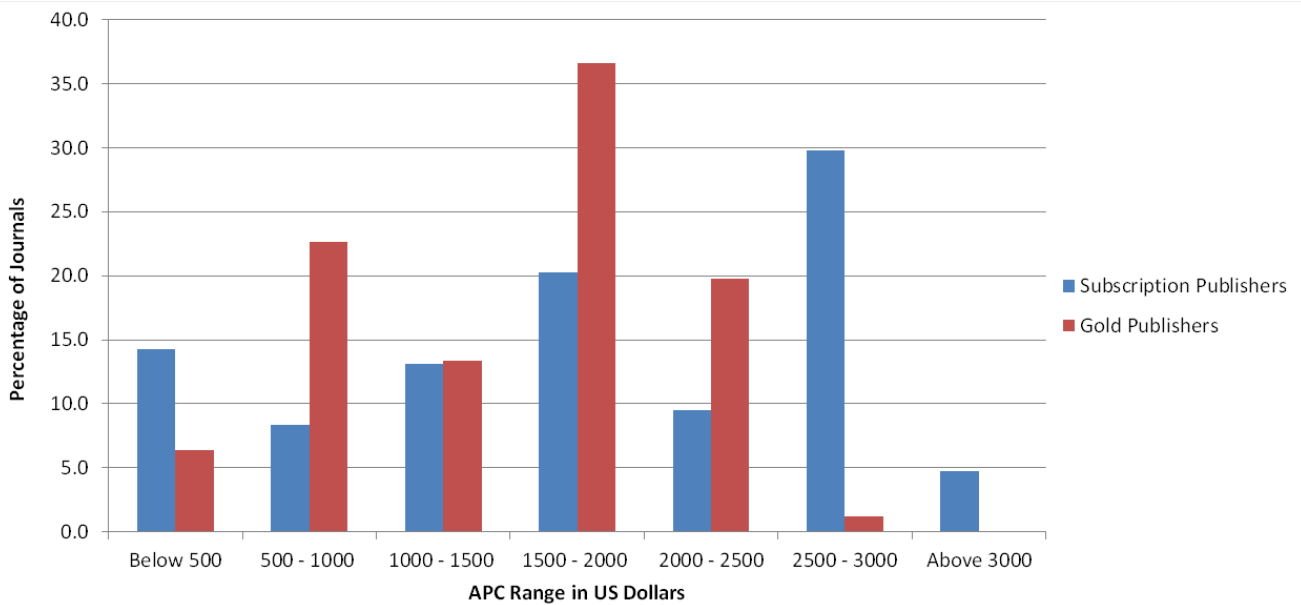


Figure 9: Distribution of APC Prices for Traditionally Subscription and OA-Only Publishers²⁶

Hybrid Journals

Between October 2009 and August 2013 the number of subscription journals from 13 major publishers offering a hybrid option increased from 1,995 to 8,003.²⁷ Many publishers have set their hybrid APC fees at or around \$3,000. Elsevier has been an exception, individually pricing each journal’s hybrid cost. By fall 2013, 1,532 Elsevier journals had a hybrid option. Björk and Solomon were able to extract the hybrid price from the websites of 1,207 of these journals. The Scopus title list data and Source Normalized Impact per Paper (SNIP) values were merged with this data. A summary of the APC pricing by discipline (using a disciplinary scheme similar to the PIF scheme developed for that study) is provided below.

²⁶ This figure originally appears as Figure 4 in Björk and Solomon (2014) Developing an Effective Market for Open Access Article Processing Charges. Report to the Wellcome Trust.

<https://wellcome.ac.uk/sites/default/files/developing-effective-market-for-open-access-article-processing-charges-mar14.pdf>

²⁷ Ibid.

Table 22: Elsevier APC Pricing, Hybrid Journals, by Discipline

Discipline Category	Average APC in USD	APC Price Range in USD	Correlation APC with SNIP*	Number of Journals**
Arts and Humanities	\$1,452	\$750 – \$1,800	0.41	25
Biomedicine	\$2,551	\$1,100 – \$5,000	0.30	487
Business and Economics	\$1,612	\$750 – \$3,300	0.39	160
Chemistry	\$2,675	\$1,000 – \$3,750	0.32	131
Earth Sciences	\$2,631	\$1,000 – \$3,750	0.15	232
Engineering	\$2,524	\$750 – \$3,750	0.21	424
Mathematics	\$2,099	\$750 – \$3,300	0.46	81
Physics and Astronomy	\$2,479	\$1,800 – \$3,750	0.36	117
Social Sciences	\$1,835	\$750 – \$3,750	0.25	201

* Source Normalized Impact per Paper.

** Since some journals are multidisciplinary, the number of journals across disciplines totals to more than 1207, the actual number of journals.

Source: Björk and Solomon (2014) Developing an Effective Market for Open Access Article Processing Charges. Report to the Wellcome Trust.

“Flipped” Journals

Using a variety of sources, Björk and Solomon identified 41 formerly subscription journals in the Morrison dataset that have been “flipped” to full OA journals charging APCs as of the spring of 2015. The average APC for these journals after flipping was \$1,825.

Table 23: Major Publisher “Flipped” Journals

<i>APC in USD</i>			
Publisher	Mean	N	Std. Deviation
De Gruyter	\$1,356	5	309.46
Elsevier*	\$1,950	7	485.63
Nature Publishing Group	\$5,200	1	
Oxford University Press	\$2,163	3	625.81
Springer*	\$1,380	13	372.11
Taylor & Francis	\$1,032	3	451.12
Wiley	\$2,408	9	550.63
Total	\$1,825	41	829.68

*In this set, Elsevier and Springer each include the stated APC for two journals that are fully funded by SCOAP3.

Ninety percent of the journals have made the transition to OA within the last three years. While these journals represent a very small percentage of the journals published by these traditionally subscription publishers, this is a new phenomenon that may represent the beginning of a much more common strategy for transitioning business models to OA. Observing this particular set of journals provides a baseline of what APCs to expect for other journals from these publishers, given the widespread conversion assumed in this project. The \$1,825 mean APC is close to the predicted “baseline” APC from the financial model described later in the report.

Ramifications for the Model

Our project sought to combine different types of data from multiple sources in order to characterize what the likely cost of APCs would be for research intensive universities in our region. By triangulating different sources of information, we felt we could derive a more robust estimate of the likely cost of APCs for our partner institutions.

Our mapping of APC pricing data to publication output and analysis of European APC payment data produced similar average APCs for articles in full OA journals, of \$1,892 and \$1,907 respectively. A separate study of journals from major publishers that have recently ‘flipped’ to OA yielded a similar average APC of \$1,825 (although an earlier 2013 survey of average APCs from the five largest subscription publishers was somewhat higher, at \$2,097). Taken together, these data points suggest a market that has coalesced around average APC expenditures in the \$1,800-\$2,000 range for the types of universities represented in our study. These averages are higher than those found in earlier studies (such as Morrison) that look at APCs across all OA journals regardless of user uptake.

A modest amount of OA publishing (about 11%) at our partner universities is in publications that do not charge an APC. While we anticipate that such journals will continue to exist in practice, our modeling sets this to one side in order to explore the ramifications of a full conversion to APC-based OA.

There is some modest evidence of disciplinary distinctions in APC pricing, with non-STEM fields (and fields such as mathematics with limited grant funding) exhibiting somewhat lower pricing than the averages above, but the data are too limited to be reliable; moreover, our cost-per-article research (see next section) suggests that costs are more likely to vary by publisher than by discipline. We therefore discount disciplinary differences in our modeling, recognizing that such differences may emerge over time.

A correlation between APC pricing and journal quality (as measured by IF or SNIP) is also evidenced in the pricing data. We explore this correlation more fully in the “Financial Model” section of our report.

The APC data analyzed here reflect a publishing environment in which there are few market controls at play. To support their authors, many funders will pay whatever a journal charges. The model our project explores seeks to address this by designing a system in which market forces can work to restrain prices.

e. Cost-Per-Article Analysis

Rationale

For any gold open access model to be viable, it must be sustainable for stakeholders beyond the libraries. To that end, the project team sought to determine whether the model we developed provided sufficient financial latitude for publishers to perform their core functions. As part of our project plan, the team initially aimed to develop a ground-up cost model for publishing in different disciplines. We chose this approach because we believed at the time that obtaining actual cost data from publishers or through the literature would prove prohibitively challenging. We also felt that actual cost data from publishers would perhaps be unduly shaped by legacy procedures and systems they may be keen to maintain, even in the event that the industry transitioned largely to a fully APC-based model. Our belief was that a ground-up model could provide a more accurate window into the true costs of a fully APC-based model.

Methodology

Initial Analysis

In setting out to develop a ground-up cost model, the project team planned to accurately reflect the full range of expenses that are necessary to run an APC-funded journal, including salaries and benefits, editorial, technical, operations, sales and marketing, and administrative. The model was also meant to address the notion of surplus, and how these surpluses can be used to fuel growth and innovation.

Through our research, however, the process of pinning specific dollar values to components of this model proved challenging. Discussions with a number of early-stage open access publishers indicated a very wide range of costs associated with running an APC-funded journal. Variability in factors as diverse as the degree of editorial oversight, office location, and editorial management platforms (to name but a few) demonstrated the difficulty in definitively stating the startup, early stage, and mature stage costs. One publisher could, for example, elect to implement Open Journal Systems and offload final manuscript typesetting to the author. On

the other hand, another publisher might elect to license journal hosting from a commercial entity such as Atypon and pay professional copy editors to provide multiple rounds of galley proofs for author approval. By some industry counts, there are close to 100 variables for which a publisher must account.²⁸ The wide range of costs associated with each of these variables means that a ground-up cost model could defensibly run from the low hundreds of dollars to several million.

Rather than proffer a model with a range of theoretical costs so wide as to be meaningless, the project team elected to return to the notion of actual cost data, and to supplement it with a comprehensive literature review. This mechanism ultimately allows us to draw a more accurate picture of not what publishing a journal could cost, but rather what publishing a journal typically does cost, on a per paper basis. Relying on actual cost data exposes the model to the issue of legacy workflows and infrastructure, such as print costs and subscription sales mechanisms. The sum total of these costs has been roughly tallied by independent analysis and is addressed in the “Ramifications for the Model” section below.

Data Sources

As a first step to understanding what it might cost in today’s publishing environment to produce journals on a Cost-Per-Article (CPA) basis, we developed four distinct data streams: a literature review, an examination of tax documents, an analysis of the feedback provided through the ALPSP publisher survey, and direct input from specific industry sources. These were supplemented by APC data detailing the fully “gold” open access journals in which authors affiliated with partner institutions have historically chosen to publish.

The Cost-Per-Article estimates we have generated for the purposes of this project are intended to cover the total cost of publication for a given journal. This means our CPA estimates assume that research articles are subsidizing non-research article content. Letters to the editor, book reviews, news items, and other content forms all cost something to produce. We assume no APC will be paid to support publication of these materials, and that research article APCs will therefore contain a small overage charge to cover the expenses associated with these other content types.

²⁸ Anderson, Kent. “Guest Post: Kent Anderson UPDATED — 96 Things Publishers Do (2016 Edition).” The Scholarly Kitchen, February 1, 2016. <http://scholarlykitchen.sspnet.org/2016/02/01/guest-post-kent-anderson-updated-96-things-publishers-do-2016-edition/>

990 Tax Forms

Not-for-profits are required to make their IRS filings public. We used the Guidestar database to identify more than 100 societies, associations, and other organizations with journal publishing operations. Of these, the tax forms of approximately 60 organizations could be clearly interpreted and compared against paper output volumes for the year in question. While not-for-profit publishers often have pricing structures that place less emphasis on surplus than their for-profit counterparts, they share similar structural elements that shape their costs - products, functionality, vendors, audience, and so forth. This means that not-for-profits can serve as a reasonable industry proxy on the cost side of the ledger.

Journal publishing revenue was pulled from Line 9, Part I, of Form 990 (Program Service Revenue) and/or relevant sections from Part III, Line 4. Journal expenses were obtained using Line 4e, Part III, Form 990, or relevant sub-items as appropriate. The narrative information from these lines provides a clear indication of journal program activity. To as great an extent possible, publishers that co-mingled journal publishing activities with other programs and services (e.g., newsletters, conferences) were excluded from this analysis. Figures were then adjusted for inflation to 2015 dollars.

Paper counts were obtained either via the publisher's website, or from Web of Science. We attempted to focus on research articles only, but in some cases included other relevant content types (e.g., conference proceedings) as well. This is consistent with the content allocations employed in other areas of the project, such as bibliometric analysis. A full list of the publishers, tax data, and paper outputs may be found on the project's data sharing site (see Appendix J). The median Cost-Per-Article for these publishers, calculated as expenses for publishing divided by the number of published research articles, is \$2,266 in 2015 USD.

ALPSP Survey

As part of our Association of Learned & Professional Society Publishers (ALPSP) member survey, we asked publishers how they set their APC pricing. Twelve respondents indicated that their pricing is based on a cost recovery model, including indirect costs and surplus. These respondents included one non-society not-for-profit (NFP), two commercial publishers, and nine learned society or professional association NFP publishers. These publishers provided their APC pricing elsewhere in the survey (which were normalized to USD), offering direct, real world CPA data points. For the data gathered through the survey, it was necessary to zero out the surplus component so that the data was consistent with the other streams. We estimated the surplus to be 13% of publishing revenues. This percentage is consistent with figures used in a number of studies included in the literature review, as well as the Research Information

Network (RIN) report on scholarly communication²⁹ and RIN’s report for the Universities UK Open Access Co-ordination Group.³⁰ However, this estimate may be seen as conservative: using median 990 data, one could estimate the surplus margin for NFP publishers as high as 31%.

The median APC set by these twelve respondents is \$2,140 in 2015 USD. To obtain the actual CPA based on this data, we subtract 13% to remove the estimated surplus component, arriving at \$1,862.

Literature Review

We performed a thorough examination of the pertinent research and ultimately examined nine prominent and reputable studies. These analyses, authored between 2003 and 2014, address the average Cost-Per-Article (CPA) that publishers encumber to generate their products. See Appendix J for data sharing details. A number of the studies provided a range of CPAs. We documented this information as estimated low, average, and high CPAs. For the studies that only provided a single figure, we listed these as average CPAs. We then adjusted all the figures into 2015 USD. We did so by first converting to USD using 2015 conversion rates, then adjusting for inflation using US government CPI data. The median CPA identified through this review, minus the estimated 13% surplus described above, is \$2,508 in 2015 USD.

Industry Input

In addition to the above, we have received some feedback from industry sources regarding their actual costs per paper. Ubiquity Press, which supports the University of California Press’s Collabra project among others, has developed a model that sets APCs at \$500 to sustain a journal that publishes 25 or more papers per year. This number holds regardless of the discipline. A commercial OA scientific publisher (“Journal A” in the table below), not quite a startup but only a few years old, has confidentially shared on background that their full-baked CPA (total costs divided by number of papers published) is in this range as well. Other recently launched or in-development APC-supported journals also shared their calculations, models, and/or actual raw data in confidence. Frontiers, a mature open access publisher with a roster of 54 journals, provides public data about their finances. Their blog cites publishing expenses of \$9.5 million, encompassing publishing operations, editor honoraria, and general &

²⁹ “Activities, Costs and Funding Flows in the Scholarly Communications System.” Research Information Network, 2008. <http://www.rin.ac.uk/system/files/attachments/Activites-costs-flows-report.pdf>

³⁰ Research Information Network. “Monitoring the Transition to Open Access: A Report for the Universities UK Open Access Co-Ordination Group,” n.d. <http://www.researchinfonet.org/wp-content/uploads/2015/09/Full-report-FINAL-AS-PUBLISHED.pdf>

administration.³¹ During that time frame, they published 5,209 research articles. In sum, the data provided were as follows (adjusted to 2015 USD):

Table 24: Limited Industry Cost-Per-Article Data

<u>Source</u>	<u>Discipline(s)</u>	<u>Cost-Per-Article</u>
Ubiquity Press	All	\$500
Journal "A"	STM	\$500
Journal "B"	STM	\$703
Journal "C"	STM	\$1,275
Frontiers	All	\$1,826
<i>Average</i>		\$960

These data are useful insofar as they reflect real-world costs for APC-supported journals that have launched with little-to-no reliance on legacy workflows and infrastructure. They therefore provide a realistic "floor" for our Cost-Per-Article variable.

[APC Data from Authors at Partner Institutions](#)

The bibliometric data gathering process provided another useful dataset for the model. Among the information collected over the course of the project were the actual gold OA publishing activities of authors at the participating institutions. Recall that the primary goal of this project is to develop "an APC-based model that is sustainable for large North American institutions and for authors and is also viable for publishers". Therefore, it is helpful to draw from current gold OA publishing activities among our author pool to better understand the economics supporting these activities. These journals are fully OA and are already using their APCs to sustain operations. They are, as such, practical examples of an APC price point that is demonstrably viable for real-world publishers. As with the ALPSP data, a 13% surplus was estimated for these publications in Cost-Per-Article calculations. The median CPA identified through this analysis, minus the estimated 13% surplus described above, is \$1,622 in 2015 USD.

³¹ "Frontiers' Financial Commitment to Open Access Publishing." Frontiers Blog, October 13, 2015. <http://blog.frontiersin.org/2015/10/13/frontiers-financial-commitment-to-open-access-publishing/>

A summary of the multiple sources for estimating Cost-Per-Articles calculations, normalized to 2015 USD, is as follows:

Table 25: Raw Cost-Per-Article (CPA) Data Summary

Source	Median Cost-Per-Article (CPA)	Notes
Literature Review	\$2,508	Normalized to include no surplus.
990 Tax Forms	\$2,266	No surplus (990-summarized expenses for journal publishing divided by the number of published research articles)
ALPSP Survey	\$1,862	Based on \$2,140 median APC for 12 ALPSP survey respondents that indicated their APC pricing was based on a cost recovery model, <i>including indirect costs and surplus</i> . 13% removed as surplus, using RIN estimates.
APC Data from Authors at Partner Institutions	\$1,622	13% removed as surplus, using RIN estimates.
Direct Industry Input	\$960	No surplus.

Ramifications for the Model

If the limited evidence detailed in the “Industry Input” section above provides a “floor” for our Cost-Per-Article variable, the ALPSP survey responses, 990 tax data, and literature review provide additional “steps” we can use to test the viability of our overall model. Note, however, that we must first zero out the print and subscription legacy costs from these inputs to accurately sketch an APC-based publishing environment. To do so, we have applied an 11% savings. This figure is based on an estimate by Bernstein Research³² that publishers will

³² Claudio Aspesi, Andrea Rosso, and Richard Wielechowski. “Reed Elsevier: Transitioning to Open Access - Are the Cost Savings Sufficient to Protect Margins?” Bernstein Research, November 26, 2012. <http://www.richardpoynder.co.uk/OAcosts.pdf>

recognize a 10-12% net savings by switching to full OA. The net savings encompasses the discontinuation of print, reduction in customer service loads, and other workflow efficiencies, balanced against higher IT costs (due to more downloads), higher administrative burdens (for collecting APC micropayments), and the loss of advance revenues and the interest these generate.

The result of both zeroing out of legacy systems and layering the aforementioned 13% surplus margins generates a range of potential costs follows:

Table 26: Adjusted Cost-Per-Article Data Summary

Input	CPA	Adjusted to Full OA (11% Savings)	Adjusted + Surplus Margin³³
Limited Industry Input Baseline	\$960	\$960 (no adj. required)	\$1,103
APC Data from Authors at Partner Institutions	\$1,622	\$1,622 (no ad. required)	\$1,864
ALPSP Survey Median	\$1,862	\$1,657	\$1,905
990 Tax Forms Median	\$2,266	\$2,017 ³⁴	\$2,318
Literature Review Median	\$2,508	\$2,232	\$2,566

It is important to note that the surplus we are factoring in is not meant to replicate current operations for established subscription-based publishers. Rather, and consistent with the aims of the project, it is meant to project a sustainability threshold. At this level, publishers can maintain their core functionality and generate a financial cushion sufficient to both withstand market volatility and support innovation. We recognize this approach may have a ripple effect for societies that subsidize other activities via their publishing revenue. Such considerations are beyond the scope of the project.

A few additional points are worth stressing at this juncture. The Cost-Per-Article projections are meant to reflect average costs for producing peer-reviewed publications. A journal with a

³³ Surplus was estimated to be 13% of total revenue; therefore production costs (the Adjusted to Full OA column) account for 87% of total revenue. Adjusted + Surplus Margin, which represents the total revenue per paper, was calculated here to maintain those percentages.

³⁴ A very small number of the publishers in this cohort are already fully open access.

low acceptance rate that spends a good deal of its resources evaluating manuscripts it ultimately rejects, may have higher costs. That said, the purpose of generating an average Cost-Per-Article is not to fully replicate publishing workflows and operations as they exist in today's ecosystem. Rather, the intention is to identify a range of numbers that would be considered as sustainable (able to support core publishing operations, with some surplus to fuel innovations) in an APC-dominant environment. In other words, the value of the CPA is not in determining whether a specific publisher could support itself, but whether journal publishing generically can be sustained at that level.

For purposes of our modeling, the assumption is that publishing may be sustainable at a Cost-Per-Article (CPA) level as low as \$1,103, inclusive of surplus. Our confidence in sustainable CPA rises with each of the steps articulated above, with the literature review median representing an upper bounds estimate of sustainable CPA for a typical journal. Within this range of CPA estimates, the data that provides the most confidence are the real-world APCs that authors at participating institutions have already paid. This particular slice of data reflects payments made to fully gold OA journals where researchers from partner institutions actually publish. On issues of APC pricing, sustainability, and attractiveness to authors from participating institutions, this dataset provides evidence of sustainable price points that are able to attract authors.

VI. Financial Model

a. Total Cost Equation

Up until this point, our report has discussed a range of qualitative and quantitative variables germane to understanding the future viability of an APC-based journal publishing business model. In order to build a financial model that takes these factors into account, it was necessary to devise a systematic way of estimating the total cost to an institution of publishing research articles under that model. A formulaic approach was used to ensure that calculations could be made systematically and to make it possible to observe changes in outcome resulting from a change in a particular parameter.

We developed the following equation for determining the total cost to an institution:

$$\mathbf{APC}_{\text{total}} = \mathbf{PUB} \times \mathbf{PA} \times \mathbf{PR} \times \mathbf{APC}_{\text{avg}} \times (1 + \mathbf{AG})^y \times (1 + \mathbf{APCI})^y$$

In this equation, $\mathbf{APC}_{\text{total}}$ represents the total cost of APCs to an institution for a given time period (typically, a single year). \mathbf{PUB} , \mathbf{PA} , and \mathbf{PR} represent the articles whose payment the institution is responsible for in that year; $\mathbf{APC}_{\text{avg}}$ represents the average cost per article; and \mathbf{AG} and \mathbf{APCI} represent the growth factors associated with carrying the equation into the future.

In the equation's most general form, each variable is a single number (scalar), representing the entire institution. However, if more granularity is desired, each variable can be a series of numbers (vector) where each number represents a subset of the whole, such as a discipline or even a journal or article; values in the vector are multiplied separately and then summed to calculate the $\mathbf{APC}_{\text{total}}$. Each variable in the equation is defined below.

Variable Definitions

PUB, or number of publications

This variable represents the number of research articles published by institutionally-affiliated authors in a given year. This is a reasonably straightforward measure; both a scalar value and a vector of values broken down by any desired characteristics (discipline, journal, etc.) are easily obtainable from the bibliometric datasets described in the "Publishing Output Data" section. Other institutions could achieve similar calculations through datasets or advanced searching techniques in bibliometric databases such as Web of Science and Scopus.

PA, or proportion of research articles that are APC-funded

In a generalized application of our model, it is possible to envision that not all journals will convert to an APC-funded OA model. For example, as has been observed, many OA journals currently require no fees for either the author or reader. These journals are often funded through the support of a scholarly society, an external funder, an individual institution, or even the volunteer time of editors. It is reasonable to think that this business model will persist in the future, and articles in these journals will not need to be covered by APCs. This variable also allows for a more general evaluation of the publishing community's acceptance of the gold OA model; it is conceivable, for example, that some disciplines, or some individual journals, will not find it desirable or feasible to convert to open access in the future. Such an evaluation is outside of the scope of this project since we are investigating the sustainability of a journal publishing system that has completely converted to APCs, so in our modeling, we assume that **PA** = 1. However, in future analyses it may be desirable to model an environment that has only partially converted to the gold OA model; this variable builds in flexibility for such research.

PR, or proportion of articles where the institution will be responsible for payment

In the modeling process for this project, **PR** was defined as the proportion of papers for which an institutionally-affiliated author is the reprint/corresponding author on the paper. As discussed in the "Publishing Output Data" section, this is the most common mechanism in use today and is a likely way that primary responsibility for funding a given paper will be assigned in future. However, alternate models are also possible.

Note that while this equation presents **PUB** and **PR** separately, in this project these two variables were not actually calculated separately and multiplied together. Instead, by virtue of the large datasets available to us from our bibliometric partners, we were able to directly count the number of articles in each discipline, journal, or other subset that contained a reprint or corresponding author from each of our partner institutions. This strategy increases the accuracy and granularity of the calculation, but may entail more work than calculating the two variables separately. We recognize that it may not be possible for other institutions to replicate this direct calculation in future studies without access to such a robust dataset. Instead, the equation makes it possible to calculate **PUB** and **PR** separately through sampling or other available means.

APC_{avg}, or average APC for the institution

This variable represents the best prediction, or set of predictions, for the average APC across the entire institution or within each subset represented in a vector of values. Estimating APC

levels is a key variable in any analysis of future costs, and as we saw in the “APC Data” section, is also one of the most difficult to pin down. Many factors can contribute to the actual value of this variable when estimating an appropriate value or vector of values to use in the equation, particularly those enumerated immediately below.

General estimates for **APC_{avg}** can be based on the current state of APC pricing, APC payments, or the cost of publishing an article. For example, one potential projection is the average APC payment for publication in a full open access journal recorded in the various European payment databases surveyed as a part of this project. This overall average is \$1,907, and discipline-specific averages can be calculated in most areas as well (see “APC Data” section). Another strategy is to map recorded APC costs for open access articles into our existing dataset of publications by partner institution authors, to calculate what these authors likely paid for publication. The average APC predicted through this method is \$1,892, and again discipline-specific averages can be calculated in most fields (see, again, the “APC Data” section). Finally, a potential strategy for projecting APCs would be to base them off of the calculated, or surveyed, cost of publishing, including a surplus to support further development and innovation. Projected APCs using this method could range from \$1,103 to upwards of \$2,566, depending on the calculation strategy chosen (see the “Cost-Per-Article” section).

Institution-level factors can also have a significant impact on the value of **APC_{avg}**. For example, research-intensive institutions with an above-average share of high-quality output are more likely than other institutions to publish in highly selective journals, with resultantly higher APCs (see the “APC Data” and “Financial Model” sections for discussions of this correlation). Many of the partners on this project are likely to fall into this category. Similarly, an institution whose research is focused in a narrow field may experience APCs that are higher or lower than other disciplines; while our APC research was unable to draw concrete conclusions, it did suggest that there are some discipline-based patterns in APCs which could manifest themselves more clearly in a fully APC-funded world.

This variable can also be influenced by extrinsic parameters beyond the published APCs themselves. For example, outside forces such as market pressures (e.g. introducing competition into the marketplace) or agreements with publishers securing discounts for institutional authors may produce APCs that are lower than the list prices we see today. Estimates for this variable should strive to take into account these external forces and calculate how they will affect the APC. In developing our financial model (see “Financial Model” section), we concluded that competition would be introduced into the marketplace if payment were organized in particular ways and calculated journal-level predictions for **APC_{avg}** from our APC pricing data based on an analysis that we present in that section.

AG, or yearly growth in the number of articles published by the institution

This variable projects the percentage increase of articles published, by year. This value will generally be estimated based on the observed and predicted growth in output of the institution or subset of the institution being measured, as well as the publishing world as a whole. For example, we could predict publication growth in the future based on observed publication growth in our bibliometric dataset. Similarly, if an institution is growing quickly, a new professional school is being launched, or a new research center in a particular field is being built, a higher yearly rate of growth can be projected across the institution or just for one particular value in a vector. Alternately, if a particular field of research is seen to be slowing down in the world at large, a lower percentage could be chosen for **AG**.

APCI, or yearly APC inflation

APCs will inflate over time, however the exact inflation rate may be affected by the payment model under which we operate. For example, in a model in which the library or institution pays APCs regardless of cost, inflation may remain at today's above-CPI levels for journal subscriptions. However, a model in which there is increased competition for authors may realize a lower rate of inflation.

y, the year, defaulting to 0 when calculating for the current year. Total APC costs are calculated for a specific time period, normally a year. This variable supports calculation for future years so that totals can be estimated for a number of years, allowing for multi-year modeling.

Based on the growth factors **AG** and **APCI**, we can project **APC_{total}** for a year in the future. The variable **y** represents how far in the future (in years) the equation is making a projection. For example, if we wanted to use data from 2013 to make a projection for the year 2017, we would set **y** equal to 4. As may be expected, the further forward the prediction, the less reliable it may be.

Summary and Applications

Each of the variables in the total cost equation can be manipulated in a variety of ways to calculate costs under a given set of constraints or assumptions, from controls placed on APC price-setting to institution-specific expectations of growth in publications. Additionally, the equation can be applied more granularly, to generate more specific costs by discipline, department, etc. For example, how would the expansion of a particular research area at an institution affect the total cost of publishing to the institution? Through this equation, data can

be synthesized with institution-specific and subject-specific knowledge to create detailed projections and predict future costs attributable to an institution.

Specific applications of this model will be discussed in greater depth below.

b. Library APC Break-Even Point

Description

In order to better understand what each institution’s redirectable library journal expenditures are able to cover, we calculated the APC break-even point for each library: the average APC at which the library would be able to redirect all subscription funds to fully cover all publication charges within the scope of our project. This calculation will provide a valuable price level that we know the institution can support. We analyzed this cost under two different scenarios: one in which the library’s subscription budget was the sole source of funding, and one in which grant funding also plays a role. Note that a higher break-even point implies a more optimistic financial picture for the library, while a lower break-even point implies a more difficult financial picture.

Library-Only Approach

The initial approach was to identify an average APC for each of our partner institutions that would allow the library to cover all in-scope publication costs for its authors, using no more than the funds that were previously expended on the subscriptions in which those publications appeared. This process is equivalent to assuming that the library’s subscription expenditures are equal to the total cost variable **APC_{total}** in the cost equation presented in the “Total Cost Equation” section, and calculates the APC variable, **APC_{avg}**, as the break-even APC at which the equation would hold. Exact corresponding author counts and library subscription expenditures were used for each year of our data. Since we used actual data for our study period, the growth factors AG and APCI are 0 in this analysis. The resulting equation is **APC_{avg} = (Library Expenditures) / (PUB x PA x PR)**. Cumulative break-even APCs for each partner based on 2009-13 bibliometric and expenditure data are in Table 27.

Table 27: APC Break-Even Points, 2009-2013 Aggregated

Institution	Break-even APC
UC San Francisco	\$533
Harvard	\$709
UC Los Angeles	\$843
UC San Diego	\$1,104

Institution	Break-even APC
UC Davis	\$1,144
UC Berkeley	\$1,249
University of British Columbia	\$1,361
Ohio State	\$1,423
UC Irvine	\$1,691
UC Santa Barbara	\$1,862
UC Santa Cruz	\$2,081
UC Riverside	\$2,280
UC Merced	\$3,390

APC break-even points are low for universities like Harvard, UCSF, and UCLA, where researchers have a relatively high volume of publication. For these institutions, APCs would need to be in the \$500 to \$900 range to stay within current budgets. For institutions with a lower publication output, including newer or smaller campuses like UC Merced or UC Santa Cruz, average APCs above \$2,000 would be within budget. Most universities in the study fall somewhere in the middle, with break-even APCs for large partner institutions on the order of \$1,000 to \$1,500.

Taken together, the breakeven point for the entire UC system would be \$1,211. However, it may be misleading to treat these campuses as a single entity in an APC model, as subscription funding is specific to each campus.

In this analysis, Harvard is likely an outlier. Its publication volume was unusually high, possibly due to the inclusion of research affiliates from associated teaching hospitals. In calculating these break-even levels it is important to consider which authors to include. In practice some affiliated researchers would probably not be funded by the institution and policies would need to define eligibility. It may be worth noting again here that its library expenditure data also contained some gaps (although we believe those to be minimal).

Some degree of variability exists in the data underlying these calculations which could affect break-even levels:

- Corresponding authors who hold multiple institutional affiliations could approach any of their institutions for support to cover an article’s APC. As described in the “Author Focus Groups and Surveys” section, we have assumed the worst-case scenario for our analysis: if a corresponding author has an affiliation with one of our partner institutions, they will go to that institution for APC funding first. Our analysis of multiple author affiliations suggests that actual break-even price levels could be slightly higher, by a

factor of up to 10%, but for the purposes of our analysis we will continue to examine the worst-case scenario.

- Organizations and consortia such as CDL or OhioLINK that negotiate on behalf of multiple institutions often have central funds that they contribute to subscriptions, in addition to what the members institutions pay. This money may be redistributed in a number of ways or not at all. In our analysis, we have not included these external funding sources, as they are not reliably reassignable to participating libraries.
- Several of the OA journals in which authors from our partner institutions publish do not charge an APC; as described in the “APC Data” section, articles in these journals account for about 11% of the open access articles in our dataset. While it is not entirely accurate to assume that all OA publications will need funding, we chose to assume the worst-case scenario, in which all journals operate on an APC-funded model.

Library + Grant Funding Approach

An alternative approach is for the library to pay the full APC for institutionally-affiliated authors only if the author does not have funding available from a research grant. This approach would reduce the number of APCs the library needs to pay, thereby increasing the break-even points from library expenditures. In terms of the cost equation for this approach, an additional factor, **(1 - GF)**, can be applied to exclude papers that could be funded by a grant; **GF** is the percentage of papers from the institution which are grant-funded. We would then calculate the breakeven point as **APC_{avg} = (Library Expenditures) / (PUB x PA x PR x (1 - GF))**. These break-even points for each institution, again based on 2009-2013 bibliometric and expenditure data, are shown in the table below.

Table 28: APC Break-Even Points, Grant-Funded Papers Removed, 2009-2013 Aggregated

Institution	Break-even APC, Grant-funded papers removed
UC San Francisco	\$1,676
Harvard	\$1,983
UC Los Angeles	\$2,143
UC San Diego	\$3,245
UC Davis	\$3,331
Ohio State	\$3,369
University of British Columbia	\$3,569

Institution	Break-even APC, Grant-funded papers removed
UC Berkeley	\$3,625
UC Irvine	\$4,291
UC Santa Barbara	\$5,335
UC Santa Cruz	\$5,516
UC Riverside	\$6,368
UC Merced	\$8,125

Unsurprisingly, break-even points are much higher when papers acknowledging a grant are removed from the analysis. Additionally, these break-even points follow a similar pattern to the first approach, with universities that have a high volume of publication exhibiting a comparatively lower break-even point.

[Library APC Break-Even Point Discussion](#)

We can draw some conclusions about the overall financial picture for our partner institutions' libraries by comparing these APC break-even points with the data we gathered for the cost of publishing (see "Cost-Per-Article Analysis" section) and APC payments and pricing (see "APC Data" section).

Based on the break-even price points, both our Cost-Per-Article and APC research suggest that current and potential APC levels would be out of reach for most of our institutional partners if only existing library expenditures are available to cover APCs. Our Cost-Per-Article estimates suggested ranged from \$1,103 to \$2,566, inclusive of surplus. Similarly, we compared the break-even values to the APC estimates calculated in the "APC Data" section, based on payment data from European institutions and our own pricing data. These estimates -- \$1,907 for payment data in full OA journals and \$1,892 for APC pricing data mapped against actual partner institution publications -- are also higher than the break-even levels in the library-only approach for most of our partner institutions, including all three of the non-UC libraries.

Taken together, these comparisons demonstrate the low likelihood of redirected library subscription budgets alone covering the full APC costs for research-intensive institutions. This finding demonstrates the need for additional components in the financial model, as described in the "Total Cost Equation" section. By contrast, most of the break-even price levels in the library-plus-grant-funding scenario are well above cost-per-article estimates and average APCs, implying that incorporating grant funding has the potential to be sustainable for these institutions.

However, a financial model based solely on one of these two approaches would not prevent hyperinflation of APCs charged by publishers, since an institutional policy that APCs will be paid by the library would not incentivize publishers to keep their APC levels low. Libraries and institutions are not endowed with sufficient leverage to either drive APCs into the desired ranges or to control cost growth over time, perpetuating the challenges libraries face with the current subscription model.

c. Financial Model

The financial model outline here includes multiple sources of funding for APCs, including the library (with funds redirected from journal subscriptions), research grant funds, and other author-controlled discretionary funds that might be drawn from various institutional sources. The model can be reduced to simple forms, e.g., all APCs are paid from the library's journal subscription budget or from grant funds, or it can take complex forms, e.g., library subsidies up to a cap, then grant funds if available, or new discretionary research funds given to all eligible researchers at an institution. In such a "multi-payer" model where the library does not cover the entire APC, competition for funding is introduced into the publishing system, which applies new market pressure to APC pricing.

Model Rationale

The model offers several strengths that make it an attractive solution to the organization of APC funding. First, it enables libraries to redirect journal subscription budgets toward publication costs while allowing them to control overall costs by limiting subsidies. Second, the model brings research funders more solidly into the picture, without shifting the entire financial burden to research granting agencies. Research funders have an important stake in making the research they fund as widely accessible as possible, and many have adopted specific policies for this purpose, including making publication costs an allowable expense. Incorporating grant funds into the model mitigates the impact on the institution's total cost to publish, particularly the impact on the library budget.

Importantly, the model places authors in the role of thoughtful consumers when publishing their research. In order to publish in a higher-cost journal, an author might need to locate additional funds beyond a basic institutional subsidy, such as grants or discretionary research funds that can also be used for other aspects of their work such as travel, student support, equipment purchases, and so forth. The findings from our author focus groups and surveys document the price sensitivity of authors across disciplines and career stages, and strongly suggest that competition for authors could be effective in incentivizing publishers to make their APCs affordable. The economic analysis we present below explains this mechanism in greater detail.

Methodology

We developed the model in this section by first performing a general economic analysis of the relationship between article value and journal value, from the perspective of both authors and publishers. This analysis demonstrates through a series of economic equations how these factors are related to pricing behavior. We then performed a regression analysis on a subset of 78 OA journals in which authors at our institutional partners publish, in order to determine whether there was empirical evidence for a link between APCs and journal quality posited in the first analysis, using SNIP (Source Normalized Impact per Paper) as a proxy for quality. Our dataset for this analysis was selected by identifying publishers whose journals exhibit price discrimination behavior (i.e. who do not charge a single APC across all of their journals). This analysis allowed us to quantify the linkage between quality and price that we had identified earlier and provided us with several important pricing data points: a baseline APC for a journal of average quality, and a quantified value premium that we were able to apply to the full set of bibliometric data for our partner institutions in order to calculate an institution-specific average APC based on this quality distribution. Armed with these data, we then explored several models for distributing costs among different stakeholder groups described above (libraries, granting agencies, authors). Finally, we provide implementation examples for several of these models and discuss their issues and ramifications.

Economic Analysis of Journal Pricing

In the “Author Focus Groups and Surveys” section, we presented findings indicating that journal quality is the most significant factor to authors in deciding where to publish (see Figure 1). As authors experience some amount of price pressure requiring them to make conscious decisions regarding the use of funds for publication that might otherwise be directed to other activities, we expect that the APC set by publishers will settle at a value that is related to the quality of the journal as perceived by authors.

The relationship between journal quality and price can be illustrated with a simplified example. Suppose that there are two types of OA journals, high quality (H) and baseline quality (B). Readers and authors agree on this classification. Assume that entry into the baseline quality space is relatively easy, and as a consequence, pricing among B-type journals is competitive. That is, the APC charged for B-type articles is close to the cost of publishing such articles, plus some surplus for innovation. Or, $P_B = c_B$, a constant (so constant returns to scale).

In the high quality space, entry is difficult and slow. Assume that there is a single journal that can charge an APC that exceeds costs (c_H), i.e. $P_H > c_H \geq c_B$. This journal enjoys some market power. This journal need not enjoy a monopoly position to exercise market power.

Suppose that articles submitted by authors can also be of either high (H) or baseline (B) quality. Each author knows the quality of his/her article prior to submission, and journal editors can determine whether a submitted article is H-type or B-type at negligible cost.

The value (V) that an author receives from publishing in a journal reflects two contributions: 1. an article's intrinsic quality, and 2. the quality of the journal. We assume that V increases as either of these two factors increases. So, if high and low quality articles are published in the same journal, the V for the high quality article, V_i^H , will always exceed the value accruing to the baseline quality article, V_i^B (due to (1)); the subscript i denotes the common journal platform. Similarly, the V associated with articles of either type is higher when published in H-type journals (due to (2)). One possible concrete interpretation of V could be citation benefits; as we saw in our qualitative research (see the "Author Focus Groups and Surveys" section), citation benefits are hugely important to researchers.

Finally, let's assume that the H-type journal never accepts baseline quality articles for publication (this assumption can also be relaxed without much loss of generality, i.e., under reasonable conditions, the H-type journal would never find it profitable to accept low quality articles).

Given this setup, consider the behavior of the high quality journal publisher. To maximize its surplus, the H-type publisher will set the APC, P_H , so that authors of high quality articles just (weakly) prefer submitting their articles to the H-type journal (if P_H was set any higher, authors of H-type articles would submit them to a B-type journal). That is, the *net* benefit or value these authors receive from publishing in an H-type journal (value minus price) is at least as large as the net value they would receive from publishing in a B-type journal:

$$V_H^H - P_H \geq V_B^H - P_B \text{ where } V_H^H > V_B^H \quad (1)$$

In equilibrium (1) will be binding (so the two sides of (1) will be equal). Furthermore, we know that $P_B = c_B$. Making this substitution, and rearranging (1):

$$P_H = V_H^H - V_B^H + c_B \geq c_H \quad (2)$$

So, the optimal APC for the H-type journal (P_H) is increasing in two factors: (a) the "value premium" that authors of H-type articles receive by publishing in H-type journals, $V_H^H - V_B^H$, and (b) the cost of publishing a B-type article, c_B . (We assume that P_H exceeds c_H). That is, the larger the value premium or cost of publishing a B-type article, the higher is the H-type journal APC.

The fact that P_H is a function of c_B and not c_H is due to the behavioral condition expressed in (1). Since authors of H-type articles compare the net benefits of publishing in H- and B-type journals, P_H necessarily reflects V_H^H , V_B^H , and c_B ; c_H affects P_H (as a lower bound) only if the value premium and/or c_B is very low.

Given this result, suppose institutions choose to subsidize the cost of APCs for authors of B- and H-type articles. For example, suppose the subsidy is set equal to $P_B (= c_B)$. That is, authors with B-type articles face an effective price of submission equal to zero ($P_B - c_B = 0$); authors of H-type articles pay $P_H - c_B > 0$. Interestingly, this subsidy has no effect on P_H (after plugging these subsidized prices into (1), (2) is unchanged). This is because the relative net benefits of publishing in the B- and H-type journals are not affected by a subsidy level that is common to all authors.

Estimating APC Pricing

Using the economic analysis above, we are able to calculate institution-specific APC expenditures. Recall that equation 2 implies that the relationship between the APC for an article published in the H-type Journal and the value premium is

$$P_H = C_B + (V_{HH} - V_{BH}).$$

Suppose that V_{HH} can be written as the product of an article's quality, q_H , and a coefficient, β_H or $\beta_H \cdot q_H$. Similarly, $V_{BH} = \beta_B \cdot q_H$. So Equation 2 can then be rewritten as $P_H = C_B + \beta_H q_H - \beta_B q_H$. This can be simplified further:

$$P_H = \beta_1 + \beta_2 q_H$$

where $\beta_1 = C_B$ and $\beta_2 = \beta_H - \beta_B$. (2a)

Using Elsevier's Source Normalized Impact per Paper (SNIP) as a proxy for the average quality of an article published by a given journal, we estimate equation 2a using APC and SNIP data for 78 journals (the dataset is described in the "APC Data" section).^{35 36} That is,

$$APC_i = \beta_1 + \beta_2 SNIP_i$$

³⁵ Equation 2a is based on a model where only 2 levels of article quality exist. In a more general model, each additional level of quality can be represented by a similar equation. In this case the quality coefficients, β_i , increase with article quality. That is, if journal i contains higher quality content than journal j , then $\beta_i > \beta_j$. Regression results (not reported here) are consistent with this prediction. Nevertheless, using regression results based on 2a to calculate institution-specific APC costs simplifies the discussion and subsequent analysis, with little (or no expected) impact on our conclusions.

³⁶ To account for possible firm fixed effects, OA titles in the dataset had to share a publisher with at least one other OA journal. Second, since equilibrium behavior is expected to generate a positive relationship between quality and price, we only include journals where this pricing strategy is currently observed.

Our regression analysis generates the following results for equation 2a:

$$APC = 1147.68 + 709.4 * SNIP$$

We can now apply this equation to any SNIP value up to the highest one in our dataset (3.207) to estimate the APC for a particular journal. For example, an average journal in a given field, with SNIP = 1, is predicted to have an APC of \$1,857.08, and we predict that publishers will charge \$709.40 more for each additional SNIP point (up to SNIP = 3.207). For journals with a SNIP above 3.207, we set the APC at \$5000, (approximately) equal to the highest charge currently observed in the marketplace.

Applying APC Pricing to Institutional Output

We apply the regression results to the bibliometric data collected for each partner institution (see the “Publishing Output Data” section), calculating an estimated APC for each article published in the sample year. Adding together the estimated APCs for each published article yields the institution’s total cost of publication for the year.³⁷

The total costs of publication for each institution are not directly comparable without the context of how many articles the institution has published. In order to get an overall sense of the APCs that libraries would be responsible for under this model, we have calculated the average APC for each institution; these values are shown in the table below. These average APC levels are in line with estimates previously discussed from our cost per article analysis, and are only moderately (about 10%-20%) higher than estimates from our APC payment data (“APC Data” section). This is consistent with our earlier finding from mapping partner publication to current OA publications that because authors at research-intensive institutions tend to publish in higher quality journals, average APCs for our partner institutions will be higher than the general average for all OA journals. Note that while these estimates are averages, this process provides us with an institution-specific distribution of APC levels (based on current APC prices and 2013 Scopus and Web of Science data).

³⁷ To restate this in terms of the cost equation discussed in the “Total Cost Equation” section, we use the APC equation above to create a journal-level vector of values for APC_{avg} . We then create a journal-level vector of values for the counts of articles the institution is responsible for paying each year (that is, for the $PUB \times PA \times PR$ portion of the equation) from our bibliometric datasets.

Because we are using actual measured data to calculate APC_{total} for each year within the time period of our study, we are not projecting into the future. Therefore, y is 0 and so the two growth factors (AG and APCI) are eliminated from the equation. Thus, we multiply piecewise each value in the journal-level vectors for APC_{avg} and for $PUB \times PA \times PR$, and sum the resulting products to calculate APC_{total} .

Table 29: Average APC by Institution, Within PIF Financial Model, 2013

Institution	Total Cost of Publication	Papers Published	APC
Harvard	\$28.49MM	12,192	\$2,337
Ohio State	\$8.80MM	4,183	\$2,105
University of British Columbia	\$9.41MM	4,532	\$2,077
University of California System	\$57.29MM	26,044	\$2,202
UC Berkeley	\$10.41MM	4,606	\$2,260
UC Davis	\$7.49MM	3,593	\$2,085
UC Irvine	\$4.60MM	2,160	\$2,129
UC Los Angeles	\$10.59MM	4,817	\$2,198
UC Merced	\$0.51MM	233	\$2,192
UC Riverside	\$2.29MM	1,070	\$2,144
UC Santa Barbara	\$3.53MM	1,568	\$2,253
UC Santa Cruz	\$1.66MM	781	\$2,130
UC San Diego	\$9.15MM	4,111	\$2,227
UC San Francisco	\$7.05MM	3,104	\$2,271

Library APC Subsidy

Authors have differing access to grants, personal research funds, and other sources of funding for publication costs. Including a library-funded subsidy in the model addresses the case where authors of high (or even baseline) quality articles do not have access to sufficient funds to pay a journal's APC.³⁸ If a subsidy is unavailable to those authors, it follows from equation (1) in the "Economic Analysis of Journal Pricing" section that the affected articles will be published in the

³⁸ There are other important reasons for a library funding role as well, such as ensuring that ongoing stewardship and preservation needs are addressed with publishers, as discussed in the "Final Conclusions and Implications" section.

“wrong” journal (i.e., H-type articles in a B-type journal) or not at all (if the corresponding net benefit is negative).

One possible approach is to offer subsidies only to authors with insufficient grant or research funds. However, monitoring who does or does not have sufficient third-party funding poses challenges that could be difficult to overcome. An alternative and simpler approach involves offering a general subsidy up to a chosen value (or perhaps a discipline-specific value, if APCs vary across disciplines) that is close to the APC charged by the competitive “baseline quality” journals. The library then pays the lesser of the actual APC or the subsidy for each paper published by an author at that library’s institution. If the APC is higher than the library subsidy, the author must find other funds.

Since any funding sources available to the author could in general be utilized for other purposes, this approach creates an incentive for authors to exercise discretion in spending these funds. This discretion is a necessary condition to constrain pricing by publishers.

There are many possibilities for choosing an appropriate library subsidy, depending on the priorities and intentions of the institution:

- A subsidy can be calculated based on estimated APC levels: the library selects a subsidy based on the journal quality level it wishes to use as a baseline (i.e., the journal quality level that the library wishes to fully fund). A reasonable selection for this baseline level is a journal whose SNIP = 1, which is, by definition, an average journal in any discipline. In our modeling, the library subsidy calculated on this baseline journal quality level is \$1,857.
- A subsidy can be based on the measured cost of publishing an article, through any of the strategies discussed in the “Cost-Per-Article Analysis” section. A benefit of this method is that the defined subsidy is tied to what publishers are believed to require to run their journals, not what the market will bear. This approach seeks to ensure that subsidies are not high enough to lead publishers to inflate their fees. A reasonable selection for this subsidy might be the \$1,864 cost per article calculated from APC pricing data described earlier in this report, although a lower floor could also be used.
- These subsidies could be further modified by subtracting a willingness-to-pay factor (WTP), representing the amount that we anticipate authors are willing to contribute out of their own discretionary funds towards APCs. A reasonable estimate for WTP is \$300, based on responses to the section of our author survey asking authors to indicate how much they would be willing to pay towards OA publishing from various sources.
- Subsidy levels could be set specific to the institution based on the library’s break-even point – that is, the amount of money available through the redirection of subscription expenditures. In this way, libraries would choose a subsidy that they know they can afford. For our partner institutions, these subsidies would range from amounts in the

\$500 - \$1000 range for Harvard and UCSF, likely leaving the authors to find significant additional funding, to amounts upwards of \$2000 for smaller UC campuses, likely covering the majority of article fees. At the higher end, a willingness-to-pay factor could be subtracted to ensure that author incentives are also brought to bear.

- Subsidies could be set based on the actual APC the author is being asked to pay, using a fractional structure. For example, libraries could pay half of every APC, or libraries could pay the full cost up to \$500 and half of the remainder of the cost beyond that.
- Finally, subsidies could be discipline-specific, to offer increased support to researchers in subjects where grant funding is scarce or decreased support to researchers in subjects where funding is plentiful, as described in the “Research Funding Expenditure Data” section.

Several of these potential library subsidy choices are illustrated in the examples below.

Calculating Cost Allocations

Once we define the parameters for distributing APCs among the various potential institutional stakeholders, we can alter the Total Cost Equation to calculate the expected allocation to each stakeholder. We do this by setting the parameters to split the APC_{avg} factor into the amount allocated to each stakeholder: libraries, grant agencies, and other discretionary funds. This process results in three parallel equations (we simplify below by assuming $y = 0$):

$$TC_{lib} = PUB \times PA \times PR \times APC_{lib}$$

$$TC_{grant} = PUB \times PA \times PR \times APC_{grant}$$

$$TC_{other} = PUB \times PA \times PR \times APC_{other}$$

These equations are constrained by the fact that the APC is fully covered between the three stakeholders, that is:

$$APC_{avg} = APC_{lib} + APC_{grant} + APC_{other}$$

As discussed in the “Total Cost Equation” section, these variables can be vectors, such that each equation is treated at a journal or article level. We undertake this calculation for each of the implementation examples below, using actual data from 2013; it is instructive to observe one such calculation in further detail:

In Example I, authors are expected to cover the entire APC with grant funds, if they are available. If grant funds are not available, the library offers an APC up to a given level, and authors are expected to use discretionary funds to cover any additional costs. The level chosen for the library subsidy in this example is \$1,119, representing the break-even point for a research-intensive institution with average research output (using 2013 figures for UC Davis – see Table 30).

For this example, we calculate APC_{avg} and each of its components at the article or journal level.³⁹ That is, for each article in 2013 with an institutionally-affiliated corresponding author, the predicted APC, X , is calculated based on the SNIP value of the journal it was published in, using the equation above. Then, X is allocated among the stakeholders based on the parameters:

- If the article acknowledges a grant:
 - $APC_{lib} = 0$
 - $APC_{grant} = X$
 - $APC_{other} = 0$
- If the article does not acknowledge a grant:
 - $APC_{lib} =$
 - $\$1,119$ if $X \geq \$1,119$
 - X if $X < \$1,119$
 - $APC_{grant} = 0$
 - $APC_{other} = X - APC_{lib}$

This allocation is repeated for every article in the Web of Science dataset and every journal in the Scopus-not-Web of Science (SNW) dataset to create a vector of allocations for the institution. Similarly, a vector of counts of corresponding-authored articles is created for **PUB**. For each Web of Science article, the count is 1, because the Web of Science data is provided at the single article level; for each journal in the SNW dataset, the count is the expected number of articles with an institutionally-affiliated corresponding author. As with previous calculations, we assume **PA** is 1 because we are modeling a fully-gold OA system, and **PR** is 1 because we are only including articles with a corresponding author at the institution.

We therefore multiply piecewise each vector of stakeholder-specific allocations, APC_{lib} , APC_{grant} , and APC_{other} , by the vector of article counts, **PUB**, and sum across all components⁴⁰ to calculate the total cost allocated to each stakeholder. For Example I below, this calculation resulted in an allocation of \$5.27MM to grant agencies, \$1.23MM to the institution's library, and \$0.98MM to other discretionary funding at the institution.

Implementation Examples

Below we present five examples of applying different potential payment structures to the model, and the resulting allocation of costs across the library (L), grant funds (G), and discretionary funds controlled by the author (D). Three types of research institutions are

³⁹ Publishing output data was at the article level for Web of Science and the journal level for Scopus, so these allocations are calculated accordingly.

⁴⁰ A dot product, in matrix notation

observed for each example with different research publication profiles. Background data for these examples are:

Table 30: Background Data for Financial Model Examples (2013 Data Only)

	Average-Output Research Institution (UC Davis)	High-Output Research Institution (Harvard)	Low-Output Small Institution (UC Santa Cruz)
Library budget available	\$4.02MM	\$8.21MM	\$1.56MM
# papers published	3,593	12,192	781
# papers published, non-grants	1,101	3,992	250
Break-even amount, all papers	\$1,119	\$674	\$2,003
Break-even amount, non-grant papers	\$3,651	\$2,058	\$6,247
Total estimated APCs	\$7.49MM	\$28.5MM	\$1.66MM
Institutional extramural research expenditure	\$560MM	\$762MM	\$125MM

Example I: Grant Funds Expended First

In the following example, authors are expected to use grant funds (G) to pay the entire cost of the APC, when available. If no grant funds are available, the library subsidy covers APC costs up to its break-even level (L). Any remaining costs must be paid from the author’s discretionary funds (D). This example would be consistent with the principles of the Compact for Open Access Equity (COPE), which proposes that “for an article based on grant-funded research, the funding agency should be responsible for payment of the publication charge, and the article would not be eligible for underwriting by the institution.”⁴¹

⁴¹ <http://www.oacompact.org/faq/#otherinstitutions>.

Table 31: Financial Model, Example I (Grant Funds Expended First)

	Average-Output Research Institution	High-Output Research Institution	Low-Output Institution
# APCs paid 100% by grants	2,492	8,200	531
G = Funding from grants	\$5.27MM	\$19.95MM	\$1.15MM
% total institutional research expenditures	0.9%	2.6%	0.9%
Library max subsidy	\$1,119	\$674	\$2,003
L = Funding from library	\$1.23MM	\$2.69MM	\$462K
# APCs paid 100% by library	0	0	108
D = Discretionary funding	\$0.98MM	\$5.84MM	\$55K
# APCs using discretionary funds	1,101	3,992	142
Total institutional payments (L+D)	\$2.22MM	\$8.53MM	\$517K
Institutional cost change	- \$1.80MM	+ \$0.32MM	- \$1.05MM
% institutional cost change	- 45%	+ 4%	- 67%

Since grants are asked to pay first in this model up to the full amount of subsidy needed, the share of APC costs allocated to grants is highest here than in subsequent examples. The library subsidy is set at the break-even level for its available journal subscription budget; however, since grants cover first, the library total is considerably less than the total available for redirection. Conversely, since most of the break-even price levels are lower than the baseline APC we calculated for the model, the share of APC costs allocated to other discretionary funds is quite high. Institutions with a low break-even price limit, like Harvard and UCSF, would see reductions in their library expenditures but would rely on grant funding for over half the total cost of publication, and on author discretionary funds for 20% of their APC funding. Conceivably, some library savings could be redirected toward discretionary funds in this model.

Example II: Library Funds Expended First

In this example, the library again subsidizes APC costs up to its break-even point but for every paper, regardless of available grant funding. Remaining costs above the subsidy must be paid by the author, from grant funds or discretionary funds available to him or her.

Table 32: Financial Model, Example II (Library Funds Expended First)

	Average-Output Research Institution	High-Output Research Institution	Low-Output Research Institution
Library max subsidy	\$1,119	\$674	\$2,003
L = Funding from Library	\$4.02MM	\$8.22MM	\$1.47MM
# APCs paid 100% by library	0	0	392
# APCs using grant funds	2,492	8,200	247
G = Funding from grants	\$2.49MM	\$14.4MM	\$141K
% total institutional research expenditures	0.4%	1.9%	0.1%
D = Discretionary funding	\$0.99MM	\$5.84MM	\$55K
# APCs using discretionary funds	1,101	3,992	142
Total institutional payments (L+D)	\$5.01MM	\$14.06MM	\$1.52MM
Institutional cost change	+ \$0.99MM	+ \$5.85MM	- \$45K
% institutional cost change	+ 25%	+ 71%	- 3%

Here again, by using the break-even price from the library's journal subscription budget for the library subsidy, each library is able to cover its subsidizing costs. However, in this model, the entire library journal subscription budget is used, whereas the allocation to grant funding is much lower. The amount of new funding required for discretionary allocations to authors who lack grants is unchanged, but the total cost to the institution is significantly higher than in the 'grants pay first' model. This example is still net positive for small Institutions with a high break-even point, but those with a low break-even APC, like Harvard and UCSF, would have to devote considerable institutional funding to cover APCs for authors.

Example III: Library Subsidy Linked to Predicted APC for Baseline Journal

In this example, libraries of every type subsidize APC costs up to \$1,557; this subsidy represents the predicted APC of a baseline-quality journal with SNIP = 1.0 (\$1,857, as described above),

minus a “willingness-to-pay” (WTP) factor of \$300, as described in the “Library APC Subsidy” section. The author must find additional funds from grants (if available) or other discretionary funds, to pay for any journal value above the baseline level.

Table 33: Financial Model, Example III (Library Subsidy Linked to Predicted APC for Baseline Journal)

	Average-Output Research Institution	High-Output Research Institution	Low-Output Research Institution
Library max subsidy	\$1,557	\$1,557	\$1,557
L = Funding from library	\$5.53MM	\$18.8MM	\$1.20MM
# APCs paid 100% by library	356	869	80
# APCs using grant funds	2,299	7,789	492
G = Funding from grants	\$1.43MM	\$7.25MM	\$0.33MM
% total institutional research expenditures	0.3%	1.0%	0.3%
# APCs using discretionary funds	938	3,534	209
D = Discretionary funding	\$0.53MM	\$2.40MM	\$0.14MM
Total institutional payments (L+D)	\$6.06MM	\$21.2MM	\$1.33MM
Institutional cost change	+ \$2.04MM	+ \$13.0MM	- \$0.23MM
% institutional cost change	+ 51%	+ 158%	- 15%

Because the subsidy of \$1,557 is higher than the break-even price levels for many libraries, we see that the costs allocated to those libraries are larger than their budgets available for redirection. High-output institutions would require a significant increase in the funds allocated to the library over current budgets, while the amount of discretionary funding allocated directly

to and/or required from authors would be less. Institutions with lower publication output would see a total institutional cost below current subscription budget levels.

Example IV: Reduced Library Subsidy Linked to Actual APC

In this example, the library subsidizes half of the APC for every paper. Beyond the subsidy, we again assume that the author uses grant funds if available, and otherwise uses discretionary funds under his or her control.

Table 34: Financial Model, Example IV (Reduced Library Subsidy Linked to Actual APC)

	Average-Output Research Institution	High-Output Research Institution	Low-Output Research Institution
Library max subsidy	Varies	Varies	Varies
L = Funding from Library	\$3.75MM	\$14.2MM	\$832K
# APCs paid 100% by library	0	0	0
# APCs using grant funds	2,492	8,200	531
G = Funding from Grants	\$2.64MM	\$9.98MM	\$573K
% total institutional research expenditures	0.5%	1.3%	0.5%
# APCs using discretionary funds	1,101	3,992	250
D = Discretionary funding	\$1.11MM	\$4.27MM	\$259K
Total institutional payments (L+D)	\$4.85MM	\$18.5MM	\$1.09MM
Institutional cost change	+ \$0.83MM	+ \$10.3MM	- \$475K
% institutional cost change	+ 21%	+ 125%	- 30%

In this scenario, because the library subsidy is significantly below the total APC, each library's ability to cover these costs is more secure. Grants cover a larger proportion of publication costs than in the other library-pays-first models, except for the highest-output institutions. Institutions such as Harvard and UCSF would see a significant increase in the cost to the

institution for both library and discretionary budgets, institutions with a more moderate output like UC Merced or Santa Cruz would see a reduced cost to the institution.

Example V: Variable Library Subsidy, Based on Available Grant Funding

In this example, the library offers a subsidy to all authors, but the subsidy level is lower for those papers for which there is an associated research grant. Here, the library subsidizes up to \$500 for papers with grant funding available, and up to \$1,557 (the subsidy described in Example III) for papers with no grant funding available.

As with previous examples, any remaining costs above the library subsidy can be paid by grant funds or other author discretionary funds; the calculations below assume that authors will choose to pay with grant funding when it is available.

Table 35: Financial Model, Example V (Variable Library Subsidy Based On Available Grant Funding)

	Average-Output Research Institution	High-Output Research Institution	Low-Output Research Institution
Library max subsidy	\$500 or \$1,557	\$500 or \$1,557	\$500 or \$1,557
L = Funding from Library	\$2.93MM	\$10.2MM	\$650K
# APCs paid 100% by library	164	458	42
# APCs using grant funds	2,492	8,200	531
G = Funding from Grants	\$4.03MM	\$15.9MM	\$880K
% total institutional research expenditures	0.7%	2.1%	0.7%
# APCs using discretionary funds	938	3,534	209
D = Discretionary funding	\$0.53MM	\$2.40MM	\$140K
Total institutional payments (L+D)	\$3.46MM	\$12.6MM	\$780K
Institutional cost change	-\$0.56MM	+\$4.4MM	-\$785K

	Average-Output Research Institution	High-Output Research Institution	Low-Output Research Institution
% institutional cost change	- 14%	+ 54%	- 50%

This example would significantly lower library costs and provide a modest reduction in overall costs to institutions of average output, although the highest-output campuses would still see an increase in institutional costs. In general, this example describes a reasonable middle-ground between Example I and Example III, in which institutions rely largely on grant funds to subsidize the overall cost of publishing while still providing a moderate level of institutional support to all authors.

d. Disciplinary Distinctions

Disciplinary distinctions are challenging to quantify due to a lack of sufficient raw data in most subjects outside of science and medicine. In fact, the equation predicting the APC of a journal based on its SNIP is also limited by the absence of APC pricing data in the social sciences and humanities: 69% of APC pricing observations used for the regression model are in Biomedical Research Disciplines, Clinical Medicine, or Life Sciences, while there are none in Social Science, Arts and Humanities, Business and Economics, or Psychiatry/Psychology. The use of SNIP as our journal value metric helps to address this data issue: SNIP is normalized by subject, and so conclusions that we draw from our limited dataset can more readily be applied to subjects outside of the dataset; a SNIP of 1 represents the average journal in any discipline.

However, while the SNIP addresses the disciplinary challenge on the input side of the model, it also removes any discipline specificity on the output side: we assume that an average journal has the same APC in every field. It may be reasonable to assume that the actual APCs for journals in non-science fields would settle at lower levels than journals of equal value in science and medical fields. But this assumption is not quantifiable; the ideal dataset for quantifying the relationship between APC and SNIP (as a proxy for journal value) would include several years of pricing and SNIP data, for several journals in each discipline in our PIF Subject scheme. The APC-based model is simply not mature enough outside of the sciences to provide this level of data.

A potentially effective application of disciplinary distinctions in the model would be in setting library subsidies within an institution. Just as library budget allocations often take into account differences in the cost of books and journals in different fields (or other disciplinary factors), libraries could select a baseline subsidy, and then calculate a set of discipline-specific correction factors to add to or subtract from that baseline level. The correction factors could be based on

several components, including volume and number of grants for each discipline, publishing patterns of the researchers in that discipline, and any other number of internal factors relating to institutional support for researchers. The calculation tool to be distributed as a deliverable for this project will allow for the input of a variable library subsidy based on the discipline of the published paper; the resulting cost allocation can be viewed both in total and by discipline, allowing for the observation of overall effectiveness of the discipline-specific subsidies. Iterating the processes of selecting these discipline-specific correction factors and observing the resulting cost allocations can help libraries to choose the appropriate subsidies for their institution.

Because APCs are estimated at the article or journal level, we can also divide the total cost of publication for each institution into discipline. The resulting data can be compared to discipline-level expenditure data, where available, to observe how a shift in business model from subscriptions to APC funding could create a shift in the level of support offered to researchers in each discipline. In Table 8, we present discipline-level, UC system expenditures for CDL-negotiated content packages; below we compare these with estimated discipline-level expenditures under the APC-funded model. The relative levels of support needed, by discipline, are similar, with potentially more support needed for Clinical Medicine and less needed for Chemistry, Life Sciences, and Physics and Astronomy.

Table 36: Comparison of Subscription Expenditures to Estimated APC Expenditures, by Discipline

Subject	Percentage of CDL-negotiated package expenditures, subscription model	Percentage of total cost of publishing, APC-funded model
Arts and Humanities	4%	2%
Biomedical Research Disciplines	14%	15%
Business and Economics	3%	2%
Chemistry	11%	6%
Clinical Medicine	10%	18%
Earth Sciences	7%	6%
Engineering	12%	12%
Life Sciences	13%	9%
Mathematics	3%	2%
Multidisciplinary	<1%	1%
Physics and Astronomy	12%	9%
Psychiatry/Psychology	3%	4%
Social Science	7%	6%

Subject	Percentage of CDL-negotiated package expenditures, subscription model	Percentage of total cost of publishing, APC-funded model
Out of Scope	2%	N/A
No PIF Category	N/A	7%
Total	100%	100%

e. Grant Fund Tracking

While grant funding provides an obvious and practical solution to increased publication costs for large, research-intensive institutions, there are potential challenges to overcome in applying these funds to APC payments. First, some authors may be unwilling to use scarce grant funds to cover publications costs (although they frequently do so today, in many disciplines). Second, grants can expire before the papers they produce are published; no-cost extensions may be needed in order to utilize such funds in many instances. Third, it may be difficult in some cases to assign APC cost responsibility for a specific paper to a specific grant, especially when multiple grants cover a large research project. Finally, institutions may currently lack the necessary infrastructure to determine whether authors have grant funding that might be applied to APCs.

With respect to grant funding verification, a number of infrastructure efforts are underway to establish better linkages between grants and research papers. For example, the FundRef identifiers in article-level metadata for publications that are indexed in Web of Science, Scopus, PubMed, Google Scholar, and other easily searchable publication databases, should make verification of grant status easier in the future. On the whole, the mechanics of applying grant funds to APC payments are worthy of further investigation.

f. Model Viability

The equations we have developed allow us to calculate overall publication costs based on a specific set of institutional attributes, but they do not tell us how to organize funding or optimize costs. Considering the examples above, for how publication funding might be organized within an institution (that is, who should pay), and in order to achieve viability and sustainability for institutions, authors, and publishers in an entirely APC-funded journal publishing system, we suggest the following:

1. *Additional funding sources, including but not limited to grant funding, may supplement library serials budgets to pay for open access publication charges.* As discussed in the "Financial Model" section, our analysis confirmed that for larger research-intensive institutions, publication charges in a fully APC-based OA environment are likely to

exceed current journals budgets alone. Additional funds available to the researcher, including grant funding, should be considered to "top off" the funds redirected from libraries.

2. *Libraries have an essential funding role in any scenario.* As the entity responsible for managing subscription fees today, libraries are well-positioned and experienced in managing publication costs. More importantly, the library mission alone within the academy is dedicated to ensuring long-term stewardship and preservation of the scholarly record, functions which will continue to be critical in an open access environment. While responsibility for funding could be shifted elsewhere within the university, libraries are the organizational unit best equipped to exercise the institution's fiduciary role in managing research publication as a long-term asset. Our author surveys confirmed that faculty trust the library to act in their best interests in overseeing the financing and administration of open access.
3. *Establishing the right marketplace incentives is a key component of any funding model.* These incentives should be designed to encourage competition and restrain costs.
4. *To achieve a functional incentive structure, authors need to have "some skin in the game".* A key reason for the dysfunction of the current journals system is the role of libraries as sole financial intermediary. In academic publishing, journals compete for authors, and authors alone can exercise choice in deciding where to publish. Our surveys strongly indicated that authors would be price-sensitive in any publication decisions. Author choice is thus an important source of market discipline.
5. *Authors should not bear an undue burden in an APC-driven model.* While author participation is critical, the system should be carefully designed so as not to be financially or administratively burdensome.

In considering the viability of the financial model, several factors must be taken into account. At the most basic level, we can compare the library's APC cost allocation to redirecting its journal subscription budgets and other OA-related payments. Careful selection of the library subsidy can ensure that the costs allocated to the library stay below budget limits (in this case, the expenditures available for redirection) and meet the requirement that APC costs not exceed current budgets.

However, the portion allocated to the library is only one component of the APC, and the other allocations must also be viable. The portion of costs allocated to grant funds can be set at various levels, depending on whether grants or institutional funding sources are asked to pay

first. These levels can be compared to the data on extramural research expenditures, gathered in the “Research Funding Expenditure Data” section; this comparison appears below in 2015 USD. For most institutions, costs allocated to grants represent a very low percentage of the total research expenditures for the institution. However, even in a library-pays-first model, institutions with a low per article library subsidy (such as Harvard or UCSF) would be particularly reliant on grant funds to cover publishing costs. Note that since the Institutional Extramural Expenditures include indirect cost recovery, the percentages in the tables below represent the percentage of all research expenditures at the institution, and not just a percentage of the expenditures under the control of researchers.

Additionally, a modest percentage of authors in our study (17%) felt that it would be inappropriate to utilize grant funds to pay open access charges. Although a minority, this finding suggests that a culture shift will be necessary for some authors to adopt the practice of using grant funds for APCs and to make this a standard line item in grant proposals. For some, the total allocation to grant funds may be more than those authors are willing to earmark for APCs in their grant budgets.

Table 37: Allocation from Extramural Grant Funds, in Actual Dollars and as a Percentage of the Institution’s Total Research Expenditures

Institution	Institutional Extramural Research Expenditures (thousands of US dollars)	Example I Grant Pays First (thousands of US dollars)		Example II Library Subsidy = Break-Even Point (thousands of US dollars)		Example III Library Subsidy = Baseline APC – WTP (thousands of US dollars)	
Harvard	\$762,000	\$19,950	2.62%	\$14,430	1.89%	\$7,250	0.95%
Ohio State	\$693,000	\$5,260	0.76%	\$2,020	0.29%	\$1,480	0.21%
University of BC	\$531,000	\$6,210	1.17%	\$2,200	0.41%	\$1,710	0.32%
UC System	\$4,508,000	\$39,996	0.89%	\$19,785	0.44%	\$12,674	0.28%
UC Berkeley	\$583,000	\$7,480	1.28%	\$3,500	0.60%	\$2,520	0.43%
UC Davis	\$560,000	\$5,270	0.94%	\$2,490	0.44%	\$1,430	0.26%
UC Irvine	\$296,000	\$3,060	1.03%	\$826	0.28%	\$886	0.30%
UC Los Angeles	\$803,000	\$6,940	0.86%	\$4,290	0.53%	\$2,200	0.27%

Institution	Institutional Extramural Research Expenditures (thousands of US dollars)	Example I Grant Pays First (thousands of US dollars)		Example II Library Subsidy = Break-Even Point (thousands of US dollars)		Example III Library Subsidy = Baseline APC – WTP (thousands of US dollars)	
UC Merced	\$14,100	\$316	2.24%	\$18	0.13%	\$94	0.67%
UC Riverside	\$95,600	\$1,580	1.65%	\$214	0.22%	\$482	0.50%
UC Santa Barbara	\$203,000	\$2,490	1.23%	\$616	0.30%	\$842	0.41%
UC Santa Cruz	\$125,000	\$1,150	0.92%	\$141	0.11%	\$330	0.26%
UC San Diego	\$937,000	\$6,490	0.69%	\$3,560	0.38%	\$2,090	0.22%
UC San Francisco	\$891,000	\$5,220	0.59%	\$4,130	0.46%	\$1,800	0.20%

Finally, we consider the portion of total costs allocated to other discretionary funds controlled by the author. Discretionary funds include all available non-grant sources available to authors, such as research accounts, departmental funds, and even personal funds. In the model, we considered these to be institutional funds. Many possible structures within the institution could be established to offer support to authors for this portion of funding. For example, authors could receive access to a personal discretionary research fund to use for any research activity including publishing costs. There are a number of universities that already have these types of research accounts in place. The institution, or perhaps even the library itself, could contribute additional funds to these accounts, either yearly or at start-up, with the intention of supporting publication activities. However, for many of our partner campuses, this model requires the commitment of additional funds above the existing library expenditures that could be redirected. The “total institutional payments” is the sum of the APC allocation to the library, via subsidy, and to the institution, via author discretionary funds. We compare this total to the available library expenditures for redirection in the following table (2015 USD):

Table 38: Total Institutional Responsibility, as Compared to Library Funds Potentially Available for Redirection; Library Subsidy Equal to Break-Even APC (Example II)

Institution	Redirectable Library Expenditures <i>(thousands of US dollars)</i>	Total Institutional Responsibility (Non-Grant Funds) <i>(thousands of US dollars)</i>	Institutional Responsibility, as a Percentage of Redirectable Library Expenditures
Harvard	\$8,214	\$14,060	171%
Ohio State	\$5,529	\$6,780	123%
University of BC	\$6,240	\$7,212	116%
UC System	\$30,300	\$37,420	124%
UC Berkeley	\$5,679	\$6,908	122%
UC Davis	\$4,020	\$5,010	125%
UC Irvine	\$3,464	\$3,773	109%
UC Los Angeles	\$4,157	\$6,297	151%
UC Merced	\$642	\$493	77%
UC Riverside	\$2,213	\$2,081	94%
UC Santa Barbara	\$2,812	\$2,916	104%
UC Santa Cruz	\$1,565	\$1,523	97%
UC San Diego	\$4,216	\$5,591	133%
UC San Francisco	\$1,530	\$2,919	191%

For those institutions with a modest research publication output, the current library funding would be sufficient to cover both the subsidy payments and any additional author discretionary funds needed; UC Merced, Riverside, and Santa Cruz would even see cost savings. However, as we have seen previously, high-output institutions would need to add significant money to the system in the event that grant income is treated as a secondary rather than a primary source of funding. Harvard, UCSF, and UCLA would all need to spend over 50% more than their current available library funds to ensure that the discretionary funds given to authors were sufficient to cover the costs allocated to them in the model.

Both Examples I and V, however, offer a more affordable picture for high-output institutions.

Table 39: Total Institutional Responsibility, as Compared to Library Funds Potentially Available for Redirection; Grant Funds Expended First (Example I) and Variable Library Subsidy Depending on Grant Availability (Example V)

Institution	Redirectable Library Expenditures	Example I		Example V	
		Total Institutional Responsibility	TIR, as a Percentage of Redirectable Library Expenditures	Total Institutional Responsibility	TIR, as a Percentage of Redirectable Library Expenditures
Harvard	\$8.21MM	\$8.53MM	104%	\$12.63MM	154%
Ohio State	\$5.53MM	\$3.55MM	64%	\$4.77MM	86%
University of BC	\$6.24MM	\$3.21MM	51%	\$4.67MM	75%
UC System	\$30.30MM	\$17.31MM	57%	\$26.15MM	86%
UC Berkeley	\$5.68MM	\$2.93MM	52%	\$4.55MM	80%
UC Davis	\$4.02MM	\$2.22MM	55%	\$3.46MM	86%
UC Irvine	\$3.46MM	\$1.54MM	44%	\$2.24MM	65%
UC Los Angeles	\$4.16MM	\$3.65MM	88%	\$5.18MM	125%
UC Merced	\$0.642MM	\$0.195MM	30%	\$0.267MM	42%
UC Riverside	\$2.21MM	\$0.72MM	32%	\$1.07MM	48%
UC Santa Barbara	\$2.81MM	\$1.04MM	37%	\$1.58MM	56%
UC Santa Cruz	\$1.56MM	\$0.52MM	33%	\$0.78MM	50%
UC San Diego	\$4.22MM	\$2.67MM	63%	\$4.09MM	97%

		Example I		Example V	
Institution	Redirectable Library Expenditures	Total Institutional Responsibility	TIR, as a Percentage of Redirectable Library Expenditures	Total Institutional Responsibility	TIR, as a Percentage of Redirectable Library Expenditures
UC San Francisco	\$1.53MM	\$1.83MM	120%	\$2.93MM	191%

Note: Total Institutional Responsibility = Subsidy Allocation + Discretionary Allocation

When grants are called upon as the primary source of APC funding (Example I), institutional funding responsibility is a fraction of current expenditures (ranging from 30% to 88%) for eight of our partner institutions, with a modest increase of 4% for Harvard and a somewhat larger (20%) increase for UCSF. If a modest institutional subsidy is provided as a floor when grant income is available (Example V), the model still yields substantial savings for seven partners, although Harvard, UCSF, and UCLA would all see significant increases. This reinforces our assessment that for the most research-intensive institutions, grant funds will be a sine qua non of OA support via article processing charges, but that for the majority of institutions in this category, institutional and grant funds can be combined in more flexible ways to achieve viability.

A common concern is that providing a library subsidy could incentivize publishers to increase their APCs to the subsidy level available. One possible approach to that concern involves changing author incentives, along the lines of the discretionary funds in the financial model. If libraries provide a fixed subsidy to authors regardless of the actual APC and authors are allowed to bank the difference between the subsidy and the actual APC, e.g. in their discretionary research account, that would ensure that author price sensitivity is introduced at all journal value levels. In this scenario, low library subsidies (such as for Harvard or UCSF in Example 2 above) would not lead to significant additional costs, as these subsidies will not exceed the APC for very many papers. But libraries with higher break-even levels (such as UC Santa Cruz in Example 2) would greatly increase institutional costs if their break-even cost were used as a subsidy. In general, with a reasonably chosen subsidy, the institutional cost increase for this scenario would be modest: across our entire dataset, libraries covering \$1,557 of every APC only increases costs by about 1% over libraries covering the lesser of the APC or \$1,557.

Factoring in Growth

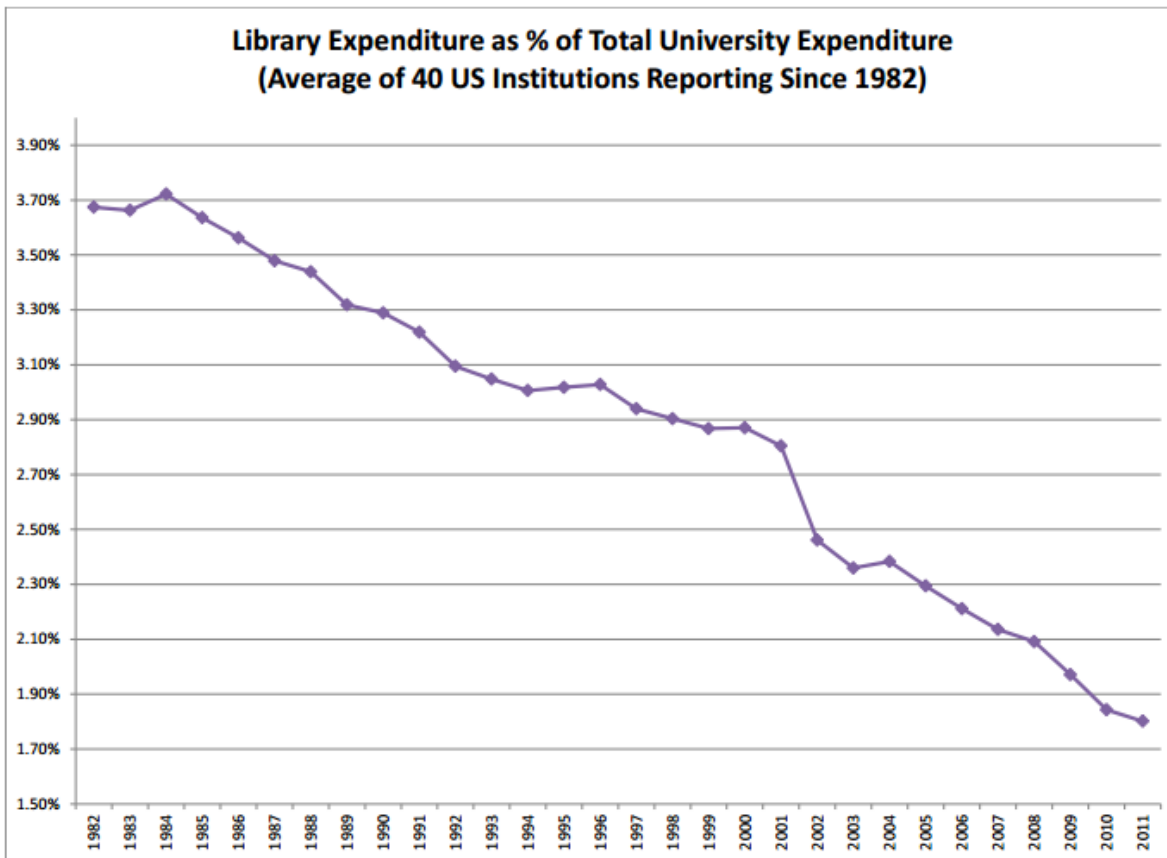
We sought five years of longitudinal data for our project in order to assess the sustainability of an APC model going forward in time. Comparing library budget trajectories to trends in publishing output should allow us to see whether a model that appears viable today will be sustainable into the future under existing conditions. While this was not explored as fully as other aspects of the model, the data points to major issues that will need to be addressed.

Our data on budget growth showed library budgets and journal expenditures that were flat or declining over the five-year period of our study (see Table 10). Within those overall figures, individual partner institutions fared better or worse during this time period. Publishing output, by contrast, largely grew at an independent rate; where there are declines, they are less precipitous. This variation is shown in the table below.

Table 40: Percent Change in Redirectable Library Expenditures vs. Publishing Volume, 2009-2013

	Redirectable Library Expenditures	Publishing Volume
Harvard	7.7%	15.9%
Ohio State	-1.1%	13.4%
University of British Columbia	11.2%	6.7%
UC Berkeley	-1.4%	2.0%
UC Davis	-3.9%	2.6%
UC Irvine	-19.8%	-5.5%
UC Los Angeles	-9.8%	-1.8%
UC Merced	10.9%	80.7%
UC Riverside	-23.4%	0.3%
UC Santa Barbara	-14.8%	-6.7%
UC Santa Cruz	-18.3%	-4.4%
UC San Diego	-10.9%	8.5%
UC San Francisco	-8.3%	12.0%

We know this lack of congruence to be true on a macro scale as well; in the US, research library budgets have been declining as a percentage of university expenditure at least since 1982, while publication output, tracking growth in research expenditure, has increased at a steady rate of upwards of 3% per year over a very long timeframe, much of it driven by growth in researchers and research expenditures.



© Association of Research Libraries, 2013

Figure 10: ARL Library Expenditures as % of Total University Expenditure, 1982-2011⁴²

⁴² http://www.arl.org/storage/documents/eg_2.pdf

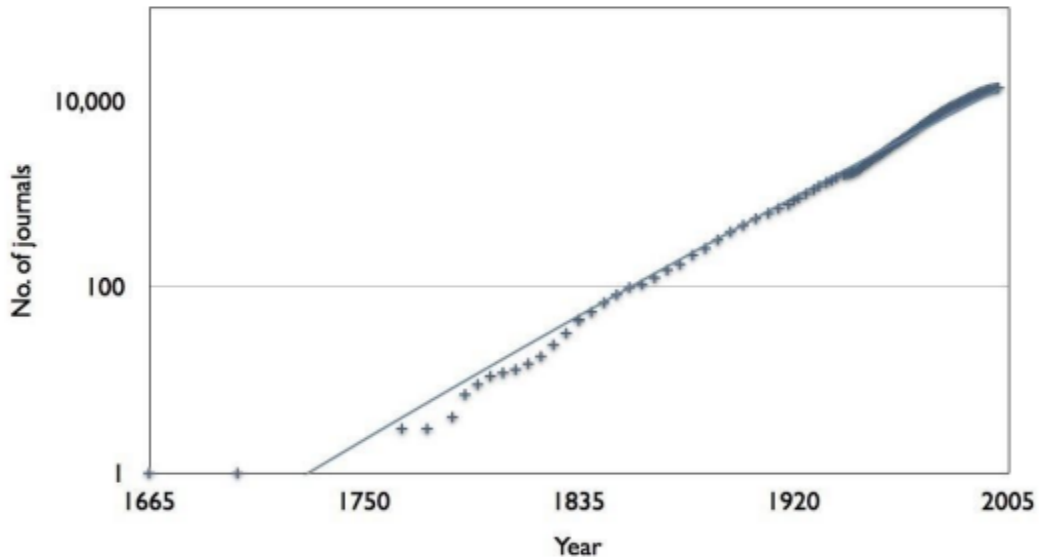


Figure 11 - The Growth of Active, Peer-Reviewed Learned Journals since 1665 (Mabe 2003)⁴³

It stands to reason that this would be so, since in the journal subscription environment, as we have seen, there is no economic relationship between library funding and article publication. Authors can continue to publish in journals at an ever-accelerating rate, even as their libraries cancel subscriptions to those very same journals.

For an APC-funded journal system to be sustainable at an institutional level at the current rate of growth in publication volume, funding capacity and support for publication output will have to come into better alignment. In the current environment, it is not necessary for an institution to connect its library funding to the publishing output of its faculty and other authors. As well, as our data have shown, a high percentage of research publication emanates from sponsored research in many of the highest-output fields. It is logical to assume that the ebb and flow of research dollars will be correlated with changes in research output.

⁴³ Ware, M, and Michael Mabe, The STM Report: An overview of scientific and scholarly journal publishing, International Association of Scientific, Technical and Medical Publishers, November 2015, p. 29 (Fig. 5). http://www.stm-assoc.org/2015_02_20_STM_Report_2015.pdf

VII. Final Conclusions and Implications

a. Implications for Long-Term Costs

Overarching Factors Affecting APC Levels

The project investigated both actual costs of publication (e.g., the cost per article) and current APC levels and found that average APCs are likely to remain in the \$1,500-\$2,000 range,⁴⁴ with exceptions, and varying in proportion to a journal's perceived quality by researchers. We posit elsewhere in this report that disciplinary differences in APCs may emerge, largely in response to differences in the availability of grant funds to subvent publication. However, it is equally possible that average APCs will not differ substantially across disciplines since publication costs -- a major factor in setting APCs -- do not vary across disciplines but rather across publishers, depending on their publication tasks, fixed costs, and surplus expectations. APC variations in the future are likely to be driven by newer publishers with lower cost structures setting lower APCs and well-established publishers charging higher APCs. Online-only publishing should also drive publication costs down, but the main driver in long-term APC levels will be author involvement in the publication payment transaction afforded by an OA business model (see below).

We are aware that some publishers at present capture revenue from non-research-producing organizations (e.g., private companies or government agencies) and that their current APC pricing does not reflect the potential loss of that revenue in an entirely OA publishing system. To maintain their current revenue, those publishers may price APCs much higher than current levels and will succeed to the extent that their journals are of high enough value to authors to continue to attract submissions.

While the project focused on the costs of a general shift to APC-funded OA journals, we are conscious of the fact that many OA journals today do not levy APCs. We are also conscious that the existence of high-quality no-fee OA journals could mitigate the financial problems that underlie this inquiry. Supporting high-quality no-fee OA journals is as desirable as supporting high-quality fee-based OA journals, however, methods to do so, and the comparative advantages of fee-based and no-fee OA journals, are beyond the scope of the present inquiry.

Current Publishers and New Platforms

Concern is also often expressed that an APC-financed publishing system will allow for further consolidation of the scholarly journal market. However, if a more open market for authors is achieved, we anticipate more competition among publishers, and their publishing platforms

⁴⁴ Note that this \$1500 to \$2000 range is referring to the APC of an average journal. In practice, our partner institutions' authors publish in above-average journals more often, and so the average APC paid by our partner institutions under this model is higher, in the \$2100 to \$2400 range (see table 30).

would contribute greatly to their competitiveness. In this scenario, the top journals that have a high degree of exclusivity could remain on traditional platforms with little consequence, but innovative new publishing platforms that have lower costs or greater benefits to authors will attract more authors and bring new players into the market. This contributes to the expectation that costs will drop over time, or at least provide greater value than today.

Author Price Sensitivity and Values

Subscription costs for scholarly journals, typically paid by research libraries, have experienced hyperinflation for decades. Containing those costs while maintaining a rigorous scholarly publishing system has so far eluded the library community, and the current system is increasingly unsustainable. If the business model for scholarly journals transitions from subscription fees to article processing charges (APCs), a reasonable question is whether APCs will also be subject to hyperinflation. The conclusion of this project, based on survey data and economic theory, is that containing future APC costs can only be achieved by involving authors more directly in the cost/benefit calculation of where to publish. As long as researchers need to publish their work but have no financial stake in that activity, there will remain very little ability to limit costs absent external regulation. Library negotiation for offsetting agreements or fixed APCs has potential to limit cost increases, but we see no reason to believe they will succeed where traditional negotiations for subscription licenses have failed. The financial model described in this report defines a framework in which libraries and their institutions continue to contribute significantly to the cost of publishing, but authors also become involved in ways that could potentially change this cost/benefit discontinuity and drive costs down over time.

Library-Publisher Relations

In the licensing-based model that is prevalent today, libraries and publishers formalize their relationships with a contract. Libraries pay publishers a set amount of money on behalf of their institution, and in turn publishers grant access to content and make a set of assurances relating to that content (terms of re-use, participation in preservation efforts, a guaranteed percentage of uptime, etc.). In this environment, libraries are able to act as a fiduciary on behalf of the institution, ensuring that proper rights are granted which protect the institution while maximizing the benefit that the institution gains from the subscription.

In the current APC-funded model, however, contracts between libraries and publishers can be extraneous; the only agreement may be the publishing agreement with the author of each paper. In such a model, libraries lose the leverage to push publishers to make choices about rights and assurances which are in the best interest of the institution. While some authors are quite knowledgeable about appropriate stewardship of published information, libraries are the organizations at the institution which have developed expertise and bear the responsibility in

this area. For the most part, authors can be expected to make choices that are in their individual best interest, but not necessarily that of the institution or the academy in general. However, libraries do have some additional options that they may explore to continue influencing the behavior of publishers and exercising their fiduciary role.

APC Contracts

Many libraries today participate in license contracts (“big deals”) with publishers, allowing their institutions to access large amounts of content at a discount off of the list price. One option for library-publisher interaction in an APC-funded system is to devise a similar contractual arrangement covering APC costs and other relevant terms. These arrangements could have notable benefits for libraries and institutions. Negotiated reductions in APCs could help libraries achieve target APC levels to fit within their allocated budgets, or to reduce the extra funds that authors would need to come up with for publication in the model. Libraries could then work into these contracts their desired stewardship terms relating to proper licensing, access provision, compliance, preservation arrangements, and so forth.

At the present time, many libraries and publishers in Europe are negotiating so-called “offsetting” arrangements in which both subscription and APCs for open access publishing are covered by a single agreement. Such agreements typically cover APC business terms and the relationship of these terms to subscription payments, standard rights and obligations, as well obligations and operational mechanics unique to the APC environment. These types of deals are generally intended as a transitional model to satisfy OA mandates and eventually lead towards a fully OA model. How to structure such arrangements is still evolving, but they can readily include a discounted pricing structure for APCs similar to the OA memberships that have already been tested with reasonable success, organize the purchase of a set of publication vouchers to distribute among faculty, or arrange a single bulk payment on behalf of the entire university based on the number of articles that the institution’s authors are expected to publish over the course of the year.

However, the degree to which these emerging arrangements can be compatible with the kind of market-driven APC business model enabled in our financial model is unclear. By negotiating APC pricing via institutional deals with publishers, libraries risk perpetuating the current situation in which publishers avoid competing for funds on the basis of author choice. While such discounts may be beneficial in the short term and may create a transitional path from subscriptions to APCs, the concept of a list-price APC determined through marketplace competition risks oblation. Moreover, because authors expect the flexibility to publish where they choose, libraries would have to negotiate agreements with large numbers of publishers. Additionally, there is some analysis which suggests that such agreements could be more susceptible to an antitrust lawsuit than existing library-publisher contracts, as they would more

clearly constrain price competition in what would otherwise be a competitive marketplace for authors (for more detail on the potential of an antitrust lawsuit, see the analysis prepared by Professor Mark McCabe in Appendix H).

Indirect Influence through Subsidy Requirements

An alternate method for libraries to exercise influence on publishers' policies relating to stewardship issues such as preservation, appropriate licensing, and related issues is to attach certain conditions to the payment or subsidy offered to authors, similar to the purchasing terms and conditions that govern many university-vendor transactions. In other words, authors must be publishing in journals that meet certain requirements in order for their APC to be subsidized or reimbursed. In this way, libraries could continue to enforce proper stewardship of information without the need to have a direct financial relationship with each publisher. Authors would not necessarily be restricted from publishing in noncompliant journals, but the unavailability of a library subsidy would likely be a sufficient incentive to keep almost all authors from publishing in journals which do not meet the criteria. In turn, the reduction in the author pool could potentially incentivize noncompliant journals to adopt policies which meet institutional requirements.

A potential challenge in this approach would be the cost of monitoring journals for appropriate policies. Libraries would need to evaluate each journal or publisher separately to determine whether they comply with the defined criteria, and authors would need to be made aware of the journal's compliance prior to submitting their manuscript. To streamline this process, libraries could collaborate to define appropriate criteria and certify publishers who meet these criteria. Authors would then only need to know that the journal is certified by this library collaboration to be comfortable submitting their manuscript, and libraries would not have to shoulder the cost of developing a publisher agreement in every case.

b. Issues Outside of Scope

Author Behavior and Academic Reward Systems

In discussions of the current high cost of journal publication and the potential to create new, lower cost journal publishing platforms, there is often a presumption that authors are flexible about where they publish their research. This project looked carefully at author behavior and preferences, as part of the qualitative research described earlier, and it is clear that authors will continue to publish in the "best" journal that they can afford, even in an entirely open access publishing system. The model for financing APC-funded journals developed by this project is based on the assumption that these values will persist for the majority of authors, and that while their publishing behavior will become more economically grounded if they are required to make economic choices about where to publish, this is not likely to lead to a wholesale shift in

publication practices. The question of why authors are unwilling to change their publishing practices is complex and well-documented elsewhere, and is attributed in part to researcher reward systems (e.g. promotion and tenure systems) that reward publication volume and journal “quality” above other possible criteria. While reforming academic reward systems could potentially incentivize authors to change their publishing behavior in more transformative ways than exist today, this issue is outside the scope of our work. Our model allows authors to publish in any journal they choose, for any reason, but requires them to have a stake in the cost of that choice, creating new financial incentives to consider alternatives – such as publishing less, or in less expensive journals, to preserve their discretionary funds for other uses.

Implications for Non-North American Research Institutions

Another question that is out of scope for this project is how well an entirely APC-funded open access publishing system would work for institutions outside of North America, particularly in the developing world where research funding is often much lower. We agree that this is an important consideration in the overall question of whether to shift the scholarly journal system to APC-funded open access, and suggest that our methodology might be helpful in answering the question. If small, less research-intensive institutions in the developing world are currently spending library funds on journal subscriptions and are willing to redirect those funds to APCs, then they would resemble some of the smaller institutions we studied for this project, and might potentially stand to benefit financially from a shift to APC-funded article publishing. This may be a fruitful area for additional research.

Implications for Other Types of Library Information Resources

An oft-quoted figure for the percentage of a library’s collections budget that goes toward journal subscriptions is approximately 70%.⁴⁵ Our data, however, show that subscription expenditures for the research journal and conference literature account for on the order of 44% (at the median)⁴⁶ of the total collections budgets of our partner institutions’ libraries. Further analysis reveals that many other serials purchases, such as topical and aggregator databases, abstract and indexing services, news and trade sources, as well as many other out-of-scope serials subscriptions also account for a significant portion of library expenditures. While scholarly journals represent a substantial portion of library collections budgets, they may not constitute as high a percentage as conventional wisdom dictates. Libraries may wish to consider how shifting to an APC-funded journal publishing system might affect available funds for these

⁴⁵ Tai Phan, Laura Hardesty, and James Hug. “Academic Libraries: 2012 First Look.” U.S. Department of Education, January 2014. <http://nces.ed.gov/pubs2014/2014038.pdf>

⁴⁶ High value is 79%, low value is 19%, but 8 of 13 partner institutions are within 5 percentage points of the median 44%.

other types of library resources, or conversely, how the need for some of these resources might change in an all-OA world.

Transition to a New APC-Funded Journal System, Including Payment Mechanics

Finally, this project looks at the question of financial sustainability and viability of a new financial model for scholarly journal publishing, but does not attempt to show how such a transition could be implemented nor how it might be operationalized. Clearly, such a transition will be extremely complex, with significant risk on many sides. Moving in this direction will require careful balancing of resources and the development of entirely new operational infrastructure (e.g., to manage the payment mechanics for APCs funded from multiple sources). While our discussion of library-publisher relations explores certain issues that gesture toward this future, further investigation will be required to map out such a transition, taking into account many differences among disciplines, institutions, research funding models, legal regimes, government policies, and much more.

Appendix A: Author Focus Group Instrument

- **Domain 1: Access to other scholars' research**
 - o *First we are going to discuss access to high quality scholarly venues*
 - o In your day-to-day work, describe how you use other scholars' research.
 - How do you currently get access to this type of research?
 - How do you feel about the level of access you have to high quality scholarly research?
 - o What do you gain from this access as an author/researcher? Teacher? Student?
 - o What kinds of constraints or obstacles arise for you in terms of access?

- **Domain 2: Participants' current publishing practices**
 - o *Next we are going to discuss your publishing practices. By publishing, we mean that something you have written has been peer reviewed and accepted as a journal article or book.*
 - o Describe the types of venues do you publish in.
 - How does this affect others' ability to access your work?
 - o What are the most important criteria you use to determine where to publish your work?
 - What parts of the research or publishing processes could change these criteria?
 - o What, if anything, do you know about article processing charges (APCs)?
 - Have you ever been asked to pay an author payment (APC- article processing charge) at the time of publishing?
 - If yes, what was your reaction? What did you think?
 - If no, what would your reaction be? Would you be willing to pay for publication?
 - Who should pay these fees? Where would you go to ask for funds?

- **Domain 3: Access to the products of participants' research**
 - o *Next we are going to talk about others accessing your research.*
 - o Describe the last time you thought about the accessibility of your research for other scholars/readers.
 - How do you think others get access to your research/publications?
 - o Is it important to you that other scholars – researchers, teachers, and students – be able to access your publications?
 - Why/why not? What does it do for you, others, or your field in general?

- **Domain 4: Opinions on Gold OA and APCs**
 - o *Next we are going to discuss different economic models of publishing.*

- Are you familiar with open access as a publishing model?
 - If so, what are your thoughts about this?
 - If not: *Gold open access is a form of publishing where the journal allows unrestricted access to its peer-reviewed scholarly research. This incurs an Article Processing Charge (paid for by author or institution) per article for the article to be published.*
 - [Give them list.] What are your thoughts about APCs?
 - How do you judge the quality of a journal?
 - Do you sense that there is a difference in quality between open access and subscription journals?
 - Is pricing related to quality? If so, how?
 - How do you think that others – including deans, department chairs, chancellors – view open access journals/APCs?
 - Suppose there was a shift in publishing models among the journals you read or publish in from subscription-based to gold OA.
 - How would this affect scholarship:
 - in your field?
 - at your institution?
 - for those accessing this research?
 - Are there other viable alternatives? Different models?
- **Domain 6: The future of publishing**
- *Our final topic area is about the future of publishing and its economic models.*
 - Which modes of publishing appeal to you as an author?
 - Which modes of publishing appeal to you as a consumer of research?
 - Do you feel that you/members of the faculty would have a say in cost models moving forward?
 - If so, how? What actions can you take/roles can you play?

Appendix B: Author Survey Instrument

You are invited to participate in a research study about the open access model of scholarly publishing. This research is part of a study funded by the Andrew W. Mellon Foundation. Every response will help us towards our goal of producing reliable knowledge about attitudes and opinions related to scholarly publishing and open access. You will also have the chance to be entered into a drawing for an iPad Mini! Other than anonymous demographic information, no sensitive items are included in the study and therefore poses no foreseeable risk. Any potentially identifying information will be removed, thus assuring that the final data set is completely anonymous. Upon publication of the results of the study, the dataset may be made publicly available through a research data repository.

You must be 18 years of age or older to participate. The survey should take no longer than 10-15 minutes to complete. All responses will be anonymous, but if you choose to participate in a drawing for an iPad Mini, then you will be asked to enter your email address at the end of the survey. This identifying information will be stored separately from responses, thereby assuring anonymity. You do not have to participate in the survey in order to participate in the drawing.

Your participation in this research is voluntary, and you may decline to participate without risk. While it is useful to be complete in your responses, you may skip any questions, and you are free to withdraw from the study at any time.

If you have any questions about the study or procedures, please contact Dr. Carol Tenopir (ctenopir@utk.edu) of the University of Tennessee. If you have questions about your rights as a participant, contact the University of Tennessee Office of Research Compliance Officer at (865) 974-7697.

By clicking on NEXT you agree you have read the above informed consent and agree to participate in this study.

First, we would like to ask you a few background questions.

[Position] Please indicate your current position.

- Faculty (1)
- Graduate Student (2)
- Postdoctoral Researcher (3)
- Other (Please specify) (4) _____

Answer If Faculty Is Selected for [Position].

[FacultyPos] Current position:

- Adjunct faculty/lecturer (1)
- Assistant Professor (2)
- Associate Professor (3)
- Professor (4)
- Research faculty (5)
- Other (Please specify) (6) _____

Answer If Faculty Is Selected for [Position].

[FacultyDegree] Highest degree:

- PhD (1)
- EdD (2)
- MD (3)
- JD (4)
- MA/MS (5)
- Other (Please specify) (6) _____

Answer If Faculty Is Selected for [Position].

[FacultyDegreeYear] Please enter the year in which your highest degree was earned:

Answer If Graduate Student Is Selected for [Position].

[StdntType] Current position:

- Masters student (1)
- PhD student (2)
- JD student (3)
- MD student (4)
- Other (Please specify) (5) _____

Answer If Graduate Student Is Selected for [Position].

[StdntDegreeYear] Anticipated year of degree:

Answer If Postdoctoral Researcher Is Selected for [Position].

[PostdocDegree] Highest degree:

- PhD (1)
- EdD (2)
- MD (3)
- JD (4)
- Other (Please specify) (5)

Answer If Postdoctoral Researcher Is Selected for [Position].

[PostdocDegreeYear] Please enter the year in which your highest degree was earned:

[AreaStudy] General area of primary scholarly activity:

- Arts and Humanities (1)
- Engineering and Computer Science (2)
- Life Sciences and Medicine (3)
- Mathematics (4)
- Physical Sciences (5)
- Social Sciences (including Business, Education, and Law) (6)
- Other (7)

[SubjDiscipline] What is your specific subject discipline? _____

Next, we would like to learn more about your behaviors and opinions related to scholarly publishing.

[NumArticle] How many journal articles or conference proceedings (as an author or co-author), including those in-press, have you published in the past 3 years?

- None (1)
- 1 - 5 (2)
- 6 - 10 (3)
- 11 - 20 (4)
- 20 or more (5)

[NumAuth] How many authors (including you) were on the last paper you published?

- No co-authors, just me (1)
- 2 (myself and one additional co-author) (2)
- 3 - 5 (3)
- 6 - 9 (4)
- 10 - 20 (5)
- 20 or more (6)

[NumBooks] How many other scholarly works (e.g., books, book chapters) have you published in the past 3 years?

- None (1)
- 1 - 5 (2)
- 6 - 10 (3)
- 11 - 20 (4)
- 20 or more (5)

[SubReposUse] Approximately what percentage of the research you have published in the last three years do you make available through the following: (Click/drag along bars below) Subject-based repositories (e.g., arXiv, PubMed Central, SSRN, RePEC, etc.):

_____ Percent: (1)

[InstReposUse] Institutional repositories (e.g., UC eScholarship, UBC cIRcle, OSU Knowledge Bank, etc.):

_____ Percent: (1)

[PersWebsiteUse] Your own personal website:

_____ Percent: (1)

[GrpsAccess] For each of the following groups, how important is it to you that they are able to access your research publications?

	Not important 1 (1)	2 (2)	3 (3)	4 (4)	Very Important 5 (5)	Not applicable (6)
Researchers/faculty at other research-intensive academic institutions (1)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Researchers/faculty at different types of academic institutions (e.g., teaching-focused) (2)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Policy-makers in government or non-government organizations (3)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Practitioners in industry and business (4)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
The general public (5)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

[GrpsAccess_other] Please list any other groups for whom you consider access to your research publications important. _____

[JourFactor] Please rate the importance of each of the following factors in choosing a journal for submission/publication of your work:

	Not important 1 (1)	2 (2)	3 (3)	4 (4)	Very important 5 (5)	Not applicable (6)
Audience (1)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Editor or editorial board (2)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Fit with scope of journal (3)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Impact factor (4)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Likelihood of acceptance (5)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Open access (6)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Quality and reputation of journal (7)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Time from submission to publication (8)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

[JourFactor_other] What (if any) other factors do you consider when choosing a journal for submission of your work? _____

Open access is a form of publishing that allows unrestricted access to peer-reviewed scholarly research. Within this model, publishers may be compensated for their efforts by the author(s) or his/her institution(s) at the point of publication rather than charging subscription fees for access to their journals.

An article processing charge (APC) is the fee that is typically paid by or on behalf of the author(s) to publish in an open access journal.

[OAFamiliar] Please rate your level of familiarity with open access publishing:

- Not at all familiar 1 (1)
- 2 (2)
- 3 (3)
- 4 (4)
- Very familiar 5 (5)

[OAPub] Have you ever published in an open access journal?

- Yes (1)
- No (2)

If No Is Selected, Then Skip To [OtherFees]

[APCPaid] Have you or your co-authors paid article processing charges (either directly or through your institution, grant, or other funds) for any of the open access articles you have published?

- Yes (1)
- No (2)

If No Is Selected, Then Skip To [OtherFees]

[NumAPC] For approximately how many articles have you paid article processing charges? _____

Answer If [NumAPC] Text Response Is Greater Than/Equal to 1

(Text) We would now like to ask about your most recent article(s) for which you paid APCs. If you have indicated that you have paid APCs for more than three articles, we will only ask you about the most recent three.

[AmtPaid_1, AmtPaid_2, AmtPaid_3] How much (in US dollars) was paid for the APCs of your most recent articles: $\$ \{ \text{Im: // Field/2} \}$? Prompted to fill in amount for article 1, 2 and 3, depending on number of articles indicated for answer.

[FundSource] What was the source(s) of these funds (check all that apply)?

- My personal funds (1)
- My lab (2)
- My co-author's personal funds (3)
- My co-authors' lab (4)
- Research funder/granting agency (5)
- A fund at my (or my co-authors') library or research office (6)
- My department (7)
- Don't recall/not sure (8)
- Other (please specify) (9) _____

Answer If None Is Not Selected for [NumArticle].

[OtherFees] Within the last 3 years, when submitting to and/or publishing in a journal, have you paid fees, other than open access fees, for (check all that apply):

- None (45)
- Submission (46)
- Color charges (47)
- Reprints (48)
- Image rights (49)
- Page charges (50)
- Other (please specify) (51) _____

If None Is Selected, Then Skip To [PubAPC]

Answer If None Is Not Selected for [NumArticle].

[FeeSources] What sources of funds have you used to pay such charges? Please check all that apply.

- My personal funds (1)
- My lab (2)
- My co-author's personal funds (3)
- My co-authors' lab (4)
- Research funder/granting agency (5)
- A fund at my (or my co-authors') library or research office (6)
- My department (7)
- Don't recall/not sure (8)
- Other (please specify) (9) _____

[PubAPC] Suppose the journals in which you typically publish became fully open access with article processing charges. If you were asked to pay an article processing charge to publish an article in one of these journals, please indicate the highest fee you would consider reasonable if the funds were coming from:

	None (1)	Less than \$100 (2)	\$100-\$499 (3)	\$500-\$999 (4)	\$1000-\$1999 (5)	\$2000-\$2999 (6)	\$3000-\$3999 (7)	\$4000 or more (8)
Your own personal funds (1)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Your discretionary research funds (2)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
An open access publication fund through the library (3)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Departmental or other institutional research funds (4)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Grant funds (5)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Other (Please specify) (6)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

[OAopinions] Please indicate your level of agreement with each of the following statements:

	Disagree Strongly 1 (1)	2 (2)	3 (3)	4 (4)	Agree Strongly 5 (5)	Not Sure (6)
Paying article processing charges for open access is a reasonable alternative to subscription fees. (1)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
In general, articles published in open access journals are of lower quality than those published in subscription based journals. (2)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
In general, the amount of an article processing charge reflects the quality of the journal. (3)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

[OAScenario] Finally, suppose the journals in which you typically publish became fully open access with article processing charges. If this were to occur, please indicate your level of agreement with each of the following scenarios:

	Disagree strongly 1 (1)	2 (2)	3 (3)	4 (4)	Agree Strongly 5 (5)	Not sure (6)
My ability to publish would be limited. (1)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
More people would read and use my research. (2)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
The overall quality of published research would increase. (3)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
People from institutions with less funding would have limited ability to publish. (4)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I would find alternative ways to publish my research. (5)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
There would be	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

increased media coverage of scholarly research. (6)						
--------------------------------------------------------------------	--	--	--	--	--	--

[Comments] Additional Comments:

After pressing *submit*, you will exit the current survey and be redirected to a separate survey where you may opt in or out of the prize drawing.

Appendix C: ALPSP Member Survey Instrument



The Association of
Learned & Professional
Society Publishers

ALPSP Member Open Access Survey

Demographics

As a reminder, your responses to all questions within this survey will be rendered anonymous so that neither you nor your organization can be identified.

* 1. Which of the following best describes your organization:

- Commercial publisher
- Learned society or professional association not-for-profit publisher
- Not-for-profit publisher (not learned society)
- University press

2. Approximately how many people (Full Time Equivalent) does your organization employ in its journal publishing operations?

3. How many journals does your organization publish that you own directly?

4. How many journals does your organization publish on behalf of partner organizations (e.g., learned societies)?

5. Please indicate the approximate percentage breakout of your journal titles (including journals you publish on behalf of partner organizations) by the following discipline(s):

Biology/Biological Sciences %	<input type="text"/>
Business %	<input type="text"/>
Chemistry %	<input type="text"/>
Computer Science %	<input type="text"/>
Earth Sciences %	<input type="text"/>
Education %	<input type="text"/>
Engineering %	<input type="text"/>
Fine/Performing Arts %	<input type="text"/>
Humanities %	<input type="text"/>
Law %	<input type="text"/>
Mathematics %	<input type="text"/>
Medicine/Health %	<input type="text"/>
Physics/Astronomy %	<input type="text"/>
Psychology %	<input type="text"/>
Social Sciences %	<input type="text"/>



The Association of
Learned & Professional
Society Publishers

ALPSP Member Open Access Survey

Current Cost Models

6. What is your **primary** journal publishing business model?

- Subscriptions
- Open Access, funded by Article Publication Charges (APCs), also known as Article Processing Charges
- Other (please specify)

7. Beyond institutional subscriptions and APCs, does your journal publishing business model rely to a nontrivial extent (more than 10% of overall journal revenue), on any of the following revenue streams? (check all that apply)

- Non-academic institutional subscriptions (e.g., corporate, government)
- Individual subscriptions
- Single-article purchases/pay-per-view
- Manuscript submission fees
- Memberships (individual or institutional)
- Advertising
- Other (please specify)

* 8. Approximately what percentage of your journals are fully open access?

* 9. Approximately what percentage of your subscription journals offer an open access option (i.e., a "hybrid" option)?

10. Please share any insights you may have about the uptake of open access options across your journal portfolio.

* 11. Do you use an APC model for **any** of your journals, including journals that are not fully open access but offer a "hybrid" option?

Yes

No



The Association of
Learned & Professional
Society Publishers

ALPSP Member Open Access Survey

Reasons for Eschewing APCs

12. Can you provide some insights into why your organization has not pursued an APC publishing model?



ALPSP Member Open Access Survey

Current Cost Models (Continued)

* 13. In 2014, approximately what percentage of your peer-reviewed articles were both open access and paid for by APCs, including but not limited to institutional membership schemes?

14. With respect to APCs, do you typically:

	Yes	No
Vary the APC price depending on the number of pages	<input type="radio"/>	<input type="radio"/>
Vary the APC price depending on the number of authors	<input type="radio"/>	<input type="radio"/>
Charge an APC for content types other than peer-reviewed articles (e.g., communications, letters to the editor, review articles)	<input type="radio"/>	<input type="radio"/>
Charge page, color, or similar charges on top of APCs	<input type="radio"/>	<input type="radio"/>

15. If you charge an APC for content types other than peer-reviewed articles (e.g., communications, letters to the editor, review articles), do you vary the APC price depending on the type of article?

- Yes
- No
- We do not charge an APC for content types other than peer-reviewed articles.

16. Recognizing that you may have offset agreements, membership schemes or waiver programs in place, please tell us the following (and include the currency you are referring to as well):

Average APC list price?

Lowest APC list price?

Highest APC list price?

* 17. When setting your APCs, do you base the price on:

- The calculated cost of publishing an article given your current cost structures, including indirect costs and profit
- Being competitive with the price points of other journals in the field
- Other (please specify)

* 18. Do you offer APC waivers for authors who would otherwise have financial difficulties paying this fee?

- Yes
- No



The Association of
Learned & Professional
Society Publishers

ALPSP Member Open Access Survey

APC Waivers

19. Approximately what percentage of authors are granted waivers?

- 1% to 10%
- 10% to 25%
- 25% to 50%
- 50% to 75%
- More than 75%

20. How does APC waiver take-up impact your overall APC pricing strategy?



The Association of
Learned & Professional
Society Publishers

ALPSP Member Open Access Survey

Pricing Planning

21. How often do you evaluate your APC pricing?

- Annually
- Every two years
- No formal schedule
- Other (please specify)

22. What factors go into your evaluation of your APC pricing strategy?

23. Do you currently offer "offset" arrangements offering credits, discounts, rebates, etc., to universities that pay both subscription charges for publications and fees to make articles open access?

- Yes
- No

24. If you do not currently offer offset arrangements, are you planning to do so in the next 1-3 years?

- Yes
- No
- Not sure



ALPSP Member Open Access Survey

Opportunities and Challenges

25. To what extent has the emergence of open access generally, and the APC model specifically, impacted the way your organization:

	No Impact	Modest Impact	Substantial Impact
Allocates financial resources across departments of the organization	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Allocates personnel across departments of the organization	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Invests research and development (R&D) resources	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Strategically plans its long-term business model	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Plans for adding new journal titles	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Plans for growing the number of articles in existing journals	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

26. How has the emergence of open access generally, and the APC model specifically, impacted how you engage with libraries?

27. Thinking about the specific questions asked in this survey, how have your perspective and business practices changed in the last 3-5 years in light of open access developments?

28. Please explain any other ways in which open access has impacted your organization.

* 29. Does your organization perceive open access generally, and the APC model specifically, as:

- Primarily a challenge to the health and future of the organization
- Primarily an opportunity to the health and future of the organization
- Neutral, on balance, to the health and future of the organization

Please explain.

30. Has your organization converted any journals from a subscription to a completely open access model (not hybrid or "open choice") to date?

- Yes
- No

If so, is there anything about the experience that you would like to share?

31. In the next 3 to 5 years, please tell us about the level of organizational reorientation that open access is likely to require.



The Association of
Learned & Professional
Society Publishers

ALPSP Member Open Access Survey

Thank You

33. Having already helped us via this survey, would you be willing to be contacted again in the future, to help us clarify our findings?

Yes

No

Thank you again for participating in this survey. For more information on the Mellon-funded University of California project, visit the [Pay-It-Forward website](#).

Appendix D: Pay It Forward (PIF) Subject Mapping

PIF Subject Category	ESI Category (Thomson Reuters)	ASJC Code (Elsevier)
Arts and Humanities	[None - assigned by presence in AHCI]	1200: Arts and Humanities
Biomedical Research Disciplines	Immunology Microbiology Molecular Biology & Genetics Neuroscience & Behavior	1300: Biochemistry, Genetics, and Molecular Biology 2400: Immunology and Microbiology 2800: Neuroscience
Business and Economics	Economics & Business	1400: Business, Management and Accounting 1800: Decision Sciences 2000: Economics, Econometrics and Finance
Chemistry	Chemistry	1500: Chemical Engineering 1600: Chemistry
Clinical Medicine	Clinical Medicine Pharmacology & Toxicology	2700: Medicine 2900: Nursing 3000: Pharmacology, Toxicology and Pharmaceuticals 3500: Dentistry 3600: Health Professions
Earth Sciences	Environment/Ecology Geosciences	1900: Earth and Planetary Sciences 2300: Environmental Science
Engineering	Computer Science Engineering Materials Science	1700: Computer Science 2100: Energy 2200: Engineering 2500: Materials Science
Life Sciences	Agricultural Sciences Biology & Biochemistry Plant & Animal Science	1100: Agricultural and Biological Sciences 3400: Veterinary
Mathematics	Mathematics	2600: Mathematics
Multidisciplinary	Multidisciplinary	1000: General
Physics and Astronomy	Physics Space Science	3100: Physics and Astronomy
Psychiatry/Psychology	Psychiatry/Psychology	3200: Psychology
Social Science	Social Sciences, general	3300: Social Sciences

Step-wise strategy for assigning a PIF Subject to Scopus source material when there is no single PIF Subject represented by more ASJC codes than the others:

- If it includes Multidisciplinary or has 4 or more subjects assigned, call it Multidisciplinary.

- If it includes Arts and Humanities, call it Arts and Humanities.
- If it includes Business and Economics, call it Business and Economics.
- If it includes Social Science, call it Social Science.
- If it includes Clinical Medicine, call it Clinical Medicine.
- If it includes Engineering, call it Engineering, unless it also includes Chemistry, in which case call it Chemistry.
- The remaining sciences, in general precedence order, are:
 - Biomedical Research Disciplines, Life Sciences, Chemistry, Earth Sciences, Physics and Astronomy, Mathematics.

Appendix E: Pay It Forward Subject Characteristics

From Web of Science data, 2009-2013:

PIF subject	ESI category (Thomson Reuters)	Pct. with grant acknowledgements	Document Growth, 2009-13	Corresponding author Percent**	Avg. authors per document
Arts and Humanities	Arts and Humanities	6%	8%	86%	1.3
Biomedical Research Disciplines	Immunology	82%	19%	45%	6.7
	Microbiology	90%	28%	50%	5.8
	Molecular Biology & Genetics	88%	23%	47%	6.5
	Neuroscience & Behavior	81%	21%	47%	5.5
Business and Economics	Economics & Business	8%	17%	57%	2.2
Chemistry	Chemistry	88%	20%	63%	4.7
Clinical Medicine	Clinical Medicine	61%	23%	46%	5.9
	Pharmacology & Toxicology	72%	24%	48%	5.4
Earth Sciences	Environment/Ecology	85%	37%	43%	4.3
	Geosciences	87%	29%	40%	4.2
Engineering	Computer Science	57%	-8%	59%	3.2
	Engineering	64%	24%	57%	3.4
	Materials Science	82%	31%	58%	4.4
Life Sciences	Agricultural Sciences	69%	21%	50%	4.5
	Biology & Biochemistry	87%	22%	54%	5.3
	Plant & Animal Science	77%	15%	50%	4.3
Mathematics	Mathematics	69%	20%	57%	2.1
Multidisciplinary	Multidisciplinary	89%	104%	49%	6.1
No Coding	No ESI coding	1%	-5%	94%	3.5
Physics and Astronomy	Physics	84%	7%	40%	8.0
	Space Science	89%	9%	26%	7.3
Psychiatry/Psychology	Psychiatry/Psychology	33%	26%	43%	3.8
Social Science	Social Sciences, general	32%	23%	55%	3.0

From Scopus Data, 2009-2013:

PIF subject	ASJC Code (Elsevier)	Document Growth, 2009-13*	Corresponding author percent*	Average authors per document
Arts and Humanities	Arts and Humanities	25%	62%	2.0
Biomedical Research Disciplines	Biochemistry Genetics and Molecular Biology	30%	50%	5.6
	Immunology and Microbiology	22%	48%	5.8
	Neuroscience	22%	50%	5.0
Business and Economics	Business Management and Accounting	15%	57%	2.2
	Decision Sciences	28%	55%	2.6
	Economics Econometrics and Finance	26%	59%	2.0
Chemistry	Chemical Engineering	31%	64%	4.3
	Chemistry	20%	60%	4.5
Clinical Medicine	Dentistry	24%	58%	4.4
	Health Professions	20%	51%	4.0
	Medicine	21%	49%	5.1
	Nursing	12%	50%	3.7
	Pharmacology Toxicology and Pharmaceutics	26%	50%	5.1
Earth Sciences	Earth and Planetary Sciences	21%	33%	4.3
	Environmental Science	34%	48%	4.0
Engineering	Computer Science	23%	58%	3.3
	Energy	47%	59%	4.0
	Engineering	41%	59%	3.8
	Materials Science	15%	56%	4.4
Life Sciences	Agricultural and Biological Sciences	36%	46%	4.5
	Veterinary	10%	52%	4.8
Mathematics	Mathematics	13%	55%	3.0
Multidisciplinary	Multidisciplinary	51%	46%	4.8
Physics and Astronomy	Physics and Astronomy	10%	39%	5.9
Psychiatry/Psychology	Psychology	26%	48%	3.2
Social Science	Social Sciences	26%	64%	2.1

Appendix F: Bibliometric Data Report Templates

Report 1: Journal Publication Data – Scopus (*submitted 5/27/15*)

- **Description:**

Basic data about the world of research articles on the whole as well as research articles published by a particular institution. Data to illuminate patterns in publishing volume, co-authorship, collaboration outside of local institutions, grant funding, and OA publishing.
- **Limits:**
 - a. Source type = Journal
 - b. Source type ≠ Trade, Trade Journal
 - c. Medline-sourced title? = No/blank
- **Facets:**
 - a. Institution (Include papers with at least one author at the listed institution):
 - i. University of British Columbia
 - ii. The Ohio State University
 - iii. Harvard University
 - iv. University of California at Berkeley
 - v. University of California at Davis
 - vi. University of California at Irvine
 - vii. University of California at Los Angeles
 - viii. University of California at Merced
 - ix. University of California at Riverside
 - x. University of California at San Diego
 - xi. University of California at San Francisco
 - xii. University of California at Santa Barbara
 - xiii. University of California at Santa Cruz
 - xiv. University of California, Office of the President
 - xv. All University of California campuses (iv through xiv)
 - xvi. Full database (no limit by institution)
 - b. Discipline
(Each of the 27 second-level disciplines, plus a set of rows including all disciplines. Articles should be counted in multiple rows if published in journals with multiple discipline categories)
 - c. Publication year
(2009, 2010, 2011, 2012, 2013)

- d. Document type
([Article or Article-in-Press], Review Article)
- **Data points:**
 - a. Count: total articles
 - b. Count: articles with one author
 - c. Count: articles with more than one author
 - d. Average: co-authors per article (i.e. all articles in data point a)
 - e. Median: co-authors per article (i.e. all articles in data point a)
 - f. Average: co-authors per article, for articles with more than one author only (i.e. all articles in data point c)
 - g. Median: co-authors per article, for articles with more than one author only (i.e. all articles in data point c)
 - h. Count: Grant-funded articles (articles with a funder acknowledgement statement, either fielded or free text)
 - i. Count: OA articles
 - j. *Count: articles with local** corresponding authors
 - k. *Count: articles with non-local corresponding authors
 - l. *Count: articles with at least one non-local co-author
 - m. *Average: non-local affiliations per article
 - n. *Median: non-local affiliations per article

*: Not applicable for sheet with no limit by institution.

**: "Local" authors are those affiliated with the institution from facet 3a currently being reported upon. "Non-local" authors are affiliated with any other institution in the world.

Report 2: Non-research article data – Scopus *(submitted 5/27/15)*

1. Description:

Data about the presence of non-research-article content and review articles in journals in particular disciplines, to inform our understanding of where this type of content resides (and help determine how to support it)

2. Limits:

- a. Source type = Journal
- b. Source type ≠ Trade, Trade Journal
- c. Publication year = 2013

3. Facets: none.

4. Data points:

- a. Source title
- b. Source print ISSN
- c. Source e-ISSN
- d. Scopus Sourcedocument ID
- e. Count: "Research articles" published (Documents of type Article, Article-in-Press, Conference Paper)
- f. Count: Review articles published (Documents of type Review Article)
- g. Count: "Non-research articles" published (Documents of all other types)

Report 3: Conference proceedings data – Scopus (*submitted 5/27/15*)

1. Description:

Basic data about the world of published conference literature, both on the whole and by institution. Data to illuminate patterns in publishing volume, co-authorship, collaboration outside of local institutions, grant funding, and OA publishing.

2. Limits:

- a. Source type = Journal, Conference Proceeding, Book, Book Series
- b. Source type ≠ Trade, Trade Journal
- c. Medline-sourced title? = No/blank
- d. Document type = Conference Paper

3. Facets:

- a. Institution (Include papers with at least one author at the listed institution):
 - i. University of British Columbia
 - ii. The Ohio State University
 - iii. Harvard University
 - iv. University of California at Berkeley
 - v. University of California at Davis
 - vi. University of California at Irvine
 - vii. University of California at Los Angeles
 - viii. University of California at Merced
 - ix. University of California at Riverside
 - x. University of California at San Diego
 - xi. University of California at San Francisco
 - xii. University of California at Santa Barbara
 - xiii. University of California at Santa Cruz

- xiv. University of California, Office of the President
- xv. All University of California campuses (iv through xiv)
- xvi. Full database (no limit by institution)
- b. Discipline
(Each of the 27 second-level disciplines, plus a set of rows including all disciplines. Papers should be counted in multiple rows if published in sources with multiple discipline categories)
- c. Publication year
(2009, 2010, 2011, 2012, 2013)

4. Data points:

- a. Count: total conference papers
- b. Count: conference papers with one author
- c. Count: conference papers with more than one author
- d. Average: co-authors per conference paper (i.e. all conference papers in data point a)
- e. Median: co-authors per conference paper (i.e. all conference papers in data point a)
- f. Average: co-authors per conference paper, for papers with more than one author only (i.e. all conference papers in data point c)
- g. Median: co-authors per conference paper, for papers with more than one author only (i.e. all conference papers in data point c)
- h. Count: Grant-funded conference papers (papers with a funder acknowledgement statement, either fielded or free text)
- i. Count: OA conference papers
- j. *Count: conference papers with local** corresponding authors
- k. *Count: conference papers with non-local corresponding authors
- l. *Count: conference papers with at least one non-local co-author
- m. *Average: non-local affiliations per conference paper
- n. *Median: non-local affiliations per conference paper

*: Not applicable for sheet with no limit by institution.

**：“Local” authors are those affiliated with the institution from facet 3a currently being reported upon. “Non-local” authors are affiliated with any other institution in the world.

Report 4: Institution publishing output – Scopus (submitted 5/27/15)

1. Description:

A summary of the publication output, by source type, of each partner institution

2. Limits:

- a. Source type = Journal, Conference Proceeding, Book, Book Series
- b. Source type ≠ Trade, Trade Journal
- c. Medline-sourced title? = No/blank
- d. Document type = Article, Article-in-Press, Conference Paper, Review Article
- e. Publication year = 2013

3. Facets:

- a. Institution (Include papers with at least one author at the listed institution):
 - i. University of British Columbia
 - ii. The Ohio State University
 - iii. Harvard University
 - iv. University of California at Berkeley
 - v. University of California at Davis
 - vi. University of California at Irvine
 - vii. University of California at Los Angeles
 - viii. University of California at Merced
 - ix. University of California at Riverside
 - x. University of California at San Diego
 - xi. University of California at San Francisco
 - xii. University of California at Santa Barbara
 - xiii. University of California at Santa Cruz
 - xiv. University of California, Office of the President
 - xv. All University of California campuses (iv through xiv above)

4. Data points:

- a. Source title
- b. Source print ISSN
- c. Source e-ISSN
- d. Source type
- e. Scopus Sourcedocument ID
- f. Count: Unique documents of type Article (one count per document with an affiliated author)
- g. Count: Unique documents of type Article-in-Press (one count per document with an affiliated author)
- h. Count: Unique documents of type Conference Paper (one count per document with an affiliated author)

- i. Count: Unique documents of type Review Article (one count per document with an affiliated author)

Report 5: Journal/Conference Publication Data – Scopus *(submitted 7/30/15)*

1. Description:

Basic data about the world of research articles and conference proceedings on the whole as well as research articles and conference proceedings published by a particular institution, organized by individual publication and year. Data at the source level (i.e. by journal, proceeding, or book series) to illuminate patterns in publishing volume, co-authorship, and collaboration outside of local institutions.

2. Limits:

- a. Source type = Journal, Conference Proceeding, Book, Book-in-Series
- b. Source type ≠ Trade, Trade Journal

3. Facets:

- a. Sourcerecord ID
- b. Year (2009, 2010, 2011, 2012, 2013)

4. General informational data for each facet:

- a. Source Title
- b. Source Type
- c. ASJC Codes

5. Data point attributes:

- a. Institution (Include papers with at least one author at the listed institution):
 - i. University of British Columbia [UBC]
 - ii. The Ohio State University [OSU]
 - iii. Harvard University [HAR]
 - iv. University of California at Berkeley [UCB]
 - v. University of California at Davis [UCD]
 - vi. University of California at Irvine [UCI]
 - vii. University of California at Los Angeles [UCLA]
 - viii. University of California at Merced [UCM]
 - ix. University of California at Riverside [UCR]
 - x. University of California at San Diego [UCSD]
 - xi. University of California at San Francisco [UCSF]
 - xii. University of California at Santa Barbara [UCSB]
 - xiii. University of California at Santa Cruz [UCSC]

- xiv. University of California, Office of the President [UCOP]
- xv. All University of California campuses (iv through xiv) [UCA]
- xvi. Full database (no limit by institution) [ALL]

- b. Document Type
 - i. Article/Article-in-Press [AR]
 - ii. Review Article [RE]
 - iii. Conference Paper [CP]

6. Data points (data for each institution/document type combination in 5a and 5b):

- a. Count: total documents
- b. Count: documents with one author
- c. Count: documents with more than one author
- d. Average: co-authors per document (i.e. all documents in data point a)
- e. Median: co-authors per document (i.e. all documents in data point a)
- f. Average: co-authors per document, for documents with more than one author only (i.e. all documents in data point c)
- g. Median: co-authors per document, for documents with more than one author only (i.e. all documents in data point c)
- h. *Count: documents with local** corresponding authors
- i. *Count: documents with non-local corresponding authors
- j. *Count: documents with at least one non-local co-author
- k. *Average: non-local affiliations per document
- l. *Median: non-local affiliations per document

*: Not applicable for aspect with no limit by institution.

**：“Local” authors are those affiliated with the institution from attribute 5a currently being reported upon. “Non-local” authors are affiliated with any other institution in the world. Note that for 5.a.xv, All UC campuses, “local” means any UC author, regardless of campus, and “non-local” means any non-UC author, regardless of campus.

Report 6: Growth of journal content over time – Scopus (submitted 7/30/15)

1. Description:

Data about changes in the publication volume of journal content, including titles added and removed from Scopus.

2. Limits:

- a. Source type = Journal
- b. Source type ≠ Trade, Trade Journal

- c. Document type = Article, Article-in-Press, Review Article, Conference Paper

3. Data points:

- a. Sourcerecord ID
- b. Source Title
- c. Source ASJC codes
- d. Coverage, by year
- e. Coverage, by volume
- f. Documents published in 2009
- g. Documents published in 2010
- h. Documents published in 2011
- i. Documents published in 2012
- j. Documents published in 2013
- k. Documents published in 2014

Report 4a: Publishing output – Web of Science (*submitted 7/30/15*)

(Note: Reports 1 through 4 were superseded by 4a through 6 for Web of Science)

- **Description:**

Data at the source title level describing the publishing output of partner libraries, as well as the type of articles that make up each source.
- **Limits:**
 - a. WoS edition = Journals (any of the three journal editions), Proceedings (either edition)
 - b. Publication year = 2013
- **Facets:**
 - a. Source abbreviation (a.k.a. “Journal 20”)
 - b. Year (2009, 2010, 2011, 2012, 2013)
- **General Informational Data:**
 - a. Source Title
 - b. Source ISSN
 - c. Source E-ISSN
 - d. Series Title (where applicable)
 - e. ESI Subject
 - f. WoS Categories (concatenated and semicolon-delimited, if possible)
 - g. WoS Edition(s)

- h. Count: “Research articles” published (Documents of type Article, Proceedings Paper, Article-Proceedings Paper, or Article-Book Chapter)
- i. Count: Review articles published (Documents of type Review)
- j. Count: “Non-research articles” published (Documents of all other types)
- **Data point attributes:**
 - a. Institution (Include papers with at least one author at the listed institution):
 - i. University of British Columbia [UBC]
 - ii. The Ohio State University [OSU]
 - iii. Harvard University [HAR]
 - iv. University of California at Berkeley [UCB]
 - v. University of California at Davis [UCD]
 - vi. University of California at Irvine [UCI]
 - vii. University of California at Los Angeles [UCLA]
 - viii. University of California at Merced [UCM]
 - ix. University of California at Riverside [UCR]
 - x. University of California at San Diego [UCSD]
 - xi. University of California at San Francisco [UCSF]
 - xii. University of California at Santa Barbara [UCSB]
 - xiii. University of California at Santa Cruz [UCSC]
 - xiv. University of California, Office of the President [UCOP]
 - xv. All University of California campuses (iv through xiv) [UCA]
- **Data points (data for each institution in 5a):**
 - a. Count: Unique documents of type Article (one count per document with an affiliated author)
 - b. Count: Unique documents of type Proceedings Paper (one count per document with an affiliated author)
 - c. Count: Unique documents of type Review (one count per document with an affiliated author)

Report 5: Journal/Conference Publication Data – Web of Science *(submitted 7/30/15)*

1. Description:

Basic data about the world of research articles on the whole as well as research articles published by a particular institution, including both journal articles and conference literature, organized by individual publication and year. Data at the source level (i.e. by

journal or proceedings paper) to illuminate patterns in publishing volume, co-authorship, collaboration outside of local institutions, grant funding, and OA publishing.

2. Limits:

- a. WoS edition = Journals (any of the three journal editions), Proceedings (either edition)

3. Facets:

- a. Source abbreviation (a.k.a. "Journal 20")
- b. Year (2009, 2010, 2011, 2012, 2013)

4. General Informational Data:

- a. Source Title
- b. Source ISSN
- c. Source E-ISSN
- d. Series Title (where applicable)
- e. ESI Subject
- f. WoS Categories (concatenated and semicolon-delimited, if possible)
- g. WoS Edition(s)

5. Data point attributes:

- a. Institution (Include papers with at least one author at the listed institution):
 - i. University of British Columbia [UBC]
 - ii. The Ohio State University [OSU]
 - iii. Harvard University [HAR]
 - iv. University of California at Berkeley [UCB]
 - v. University of California at Davis [UCD]
 - vi. University of California at Irvine [UCI]
 - vii. University of California at Los Angeles [UCLA]
 - viii. University of California at Merced [UCM]
 - ix. University of California at Riverside [UCR]
 - x. University of California at San Diego [UCSD]
 - xi. University of California at San Francisco [UCSF]
 - xii. University of California at Santa Barbara [UCSB]
 - xiii. University of California at Santa Cruz [UCSC]
 - xiv. University of California, Office of the President [UCOP]
 - xv. All University of California campuses (iv through xiv) [UCA]
 - xvi. Full database (no limit by institution) [ALL]
- b. Document type*
 - i. Article [AR]
 - ii. Review [RE]

- iii. Proceedings Paper [PP]
- iv. Article-Proceedings Paper [AP]
- v. Article-Book Chapter [AB]

6. Data points (data for each institution/document type combination in 5a and 5b):

- a. Count: total documents
- b. Count: documents with one author
- c. Count: documents with more than one author
- d. Average: co-authors per document (i.e. all documents in data point a)
- e. Median: co-authors per document (i.e. all documents in data point a)
- f. Average: co-authors per document, for documents with more than one author only (i.e. all documents in data point c)
- g. Median: co-authors per document, for documents with more than one author only (i.e. all documents in data point c)
- h. Count: Grant-funded documents (documents with a funder acknowledgement statement, either fielded or free text)
- i. Count: OA documents
- j. **Count: documents with local*** corresponding authors
- k. **Count: documents with non-local corresponding authors
- l. **Count: documents with at least one non-local co-author
- m. **Average: non-local affiliations per document
- n. **Median: non-local affiliations per document

*: Other from the dual document types specifically listed, dual document types should be included with their main category. For example, “Proceedings-Book Chapter” should be included with Proceedings; “Review-Book” should be included with Review, and “Article-Book” should be included with Article.

** : Not applicable for sheet with no limit by institution.

***: “Local” authors are those affiliated with the institution from attribute 5a currently being reported upon. “Non-local” authors are affiliated with any other institution in the world. Note that for 5.a.xv, All UC campuses, “local” means any UC author, regardless of campus, and “non-local” means any non-UC author, regardless of campus.

Report 6: Growth of journal content over time – Web of Science (submitted 7/30/15)

1. Description:

Data about changes in the publication volume of journal content, including titles added and removed from Web of Science.

2. Limits:

- a. WoS edition = Journals (any of the three journal editions)
- b. Document type = Article, Review, Proceedings Paper, any dual document types including one of these three.

3. Data points:

- a. Source abbreviation (a.k.a. "Journal 20")
- b. Source Title
- c. Source ISSN
- d. Source E-ISSN
- e. Series Title (where applicable)
- f. ESI Subject (concatenated and semicolon-delimited, if possible)
- g. WoS Edition(s)
- h. Coverage, by year
- i. Coverage, by volume
- j. Documents published in 2009
- k. Documents published in 2010
- l. Documents published in 2011
- m. Documents published in 2012
- n. Documents published in 2013
- o. Documents published in 2014

Appendix G: Library Data Request Instructions

Data requested from library partners, 6/25/15:

1. The total spending by your institution on resources included in this project, for each year from 2009 to 2013. This should be a total for each year, not an itemized list (because many publishers have confidentiality agreements that we don't want to break).
2. For each year, lists of titles and packages that you are including in this total. If you'd like to annotate the lists that I'm sending you with this email, that works, or if you'd like to generate separate lists that's fine too. The data we're looking for here are:
 - For the title lists, we'd like to have the ISSN, title, publisher, and package name if applicable.
 - For the package lists, the package name and publisher should be sufficient.
3. For each year, a division of the total spending (from #1) into how much was spent on subscriptions that are only print, only electronic, and combined print and electronic.
4. The total spending by your institution on deals negotiated by consortia in each year; this includes either payments made directly to the consortium or payments made directly to publishers where terms were negotiated by the consortium. This amount should be a subset of the total spending listed in #1 above.
5. Lists of packages that you are including in the total consortial spend for each year.
6. The total spending by your institution on memberships or subsidies to open access publishers. This represents annual fees paid to these publishers, which often result in free or discounted publishing charges; this should not include any payments made to cover publishing charges directly.
7. A list of the memberships/subsidies included in the above total.
8. A brief summary of any major changes to your library's subscriptions to journals or conference proceedings over the time period of the study (2009-2013); this includes any large package subscriptions that you may have added or canceled during that time as well as any significant journal cancellation projects.
9. Your library's total collections budget for each year in the study (2009-2013).

Appendix H: Discussion of Antitrust Potential in Large Publisher APC Agreements

Analysis prepared by Professor Mark McCabe, SKEMA Business School and Boston University

To date, “big deals”, i.e., arrangements via which institutions subscribe to bundles of a single publisher’s journals and receive a package discount in return have not yielded antitrust challenges. In speculating as to why this is so, it is useful to begin by contrasting antitrust oversight of mergers on one hand, and conspiracy/monopolization cases on the other (a Big Deal investigation, whether they involve APCs or subscriptions, would fall into the latter category). In the US, and EU, for example, there are pre-merger notification requirements for companies meeting certain financial thresholds.⁴⁷ From 2000 through 2013, 24,388 merger applications were filed in the US. The vast majority of these were approved with little or no “tire-kicking.” In only 4.4% (1,079) of the cases did the FTC or DOJ actually request additional information from the merging parties.⁴⁸ So although full-blown merger investigations are relatively rare, all mergers subject to pre-merger notification requirements must be screened, and the corresponding outcomes are public information. This screening process has resulted in several DOJ investigations (e.g. West/Thomson, Elsevier/Wolters-Kluwer) and, in the former case, a challenge to publishing mergers.

In contrast, non-merger antitrust enforcement requires active case “hunting” by the DOJ or FTC, often with support from 3rd parties, including competitors and customers hurt by the defendants’ anticompetitive behavior. This type of oversight is governed by the Sherman Antitrust Act (Section 1 is most often applied in price-fixing cases, e.g. US v. Apple (2013), Section 2 addresses monopolization, e.g. US v. Microsoft (2001)). During the 10 year period 2000-2009, the DOJ initiated 847 Section 1 cases (most of which involved price-fixing), and 37 Section 2 monopolization investigations.⁴⁹ As such, excluding price-fixing cases, the annual number of new DOJ investigations involving circumstances similar to the Big Deal (restraint of trade or monopolization) is likely fewer than 10. The reasons for this are numerous. Briefly, these cases are significantly more difficult than (price-fixing) merger investigations because the law is less settled. Second, merit-based possession of market power is not illegal; rather, red flags are raised only when market power is created and/or maintained via intentional misconduct. Both of these factors reduce the number of potential cases considerably. Of

⁴⁷ The 2016 filing thresholds under the Hart-Scott-Rodino Act may be found here: <https://www.cooley.com/2016-hart-scott-rodino-antitrust-thresholds-final> .

⁴⁸ The underlying data for these number is located at: <https://www.ftc.gov/site-information/open-government/data-sets> .

⁴⁹ Data on Section 1 and Section 2 cases may be found here: <https://www.justice.gov/sites/default/files/atr/legacy/2012/04/04/281484.pdf> .

course, for any given case, the better the legal and economic evidence, and the more unified and vocal is the set of injured parties, the better the chances are that DOJ will initiate an investigation.

In their 2004 *Antitrust Law Journal* article,⁵⁰ Aaron Edlin and Dan Rubinfeld address the merits of a Sherman Act case against subscription-based Big Deal contracts (in particular, see Section III of the paper). They argue that in contrast to a merger investigation:

Defining the relevant market is not a necessary step in a monopoly power analysis, however, if Elsevier can be shown to have already exercised substantial power over price with regard to its journals collectively.

They further suggest that a comparison of commercial and non-profit journal prices is probably sufficient to demonstrate Elsevier's exercise of market power.

However, this begs the question of whether prohibiting subscription-based Big Deal contracts would have reduced publishers' market power (via new entry), and if the benefits of Big Deal contracts (smaller institutions gaining access to more content) would be lost. Although many factors contributed to DOJ inaction over the past 10+ years, my sense is that DOJ was cognizant of publisher market power, but believed (correctly) that this market power would erode only marginally in lieu of BD contracts, and that many smaller institutions might be harmed by this change (since customer-specific pricing would be less feasible). That is, a Big Deal subscription case had a low chance of success.

In an OA environment, the antitrust landscape is likely to be far different. First, since access is open, all scholars, independent of their affiliation, will have access. Second, in lieu of any OA-based Big Deal contracts,⁵¹ and assuming that authors internalize the pecuniary costs of publication (i.e. APC are paid from their own discretionary funds), then journal platforms will need to compete as described in the "Financial Model Description" section. Of course, pricing will not be uniform. All else equal, in equilibrium, higher quality journals will be able to charge higher APCs (here journal quality is a proxy for article quality). But for journals of similar quality, any substantial difference in APCs will lower submissions to the higher priced journal(s). Similar claims can be made for differences in quality, given similar APCs, etc.

⁵⁰ Aaron S. Edlin and Daniel L. Rubinfeld. "Exclusion or Efficient Pricing: The "Big Deal" Bundling of Academic Journals" *Antitrust Law Journal* (2004)

Available at: http://works.bepress.com/aaron_edlin/37/

⁵¹ This assumes that OA Big Deals would exhibit customer-specific pricing, not at the APC level, but like a traditional subscription Big Deal, at the institutional level. Therefore, Harvard (as an example) would pay a single annual fee to Publisher X, and in exchange, Harvard authors could submit as many articles as they wish to Publisher X's journals. If, instead, only APC level contracts emerge, e.g., a fixed discount off the "retail-APC," or a flat APC charge for all articles appearing in Publisher X journals, then antitrust concerns will be diminished considerably.

A good comparison is *PLOS ONE* and Nature's *Scientific Reports*. Both are mega journals, with similar peer-review policies. *PLOS ONE* began publishing 5 years before *Scientific Reports* (2006 v. 2011), and so it benefited from being the first-mover in this "space." Currently their APCs are identical at \$1,495 (upon entry in 2011, SF matched *PLOS ONE*'s APC, then equal to \$1,350). During our sample period (2009-2013) *Scientific Reports*' impact factor was increasing, while *PLOS ONE*'s impact factor was declining (*Scientific Reports*' impact factor first exceeded *PLOS ONE*'s in 2013). An examination of our Partner data reveals that *Scientific Reports*' publication growth rate was robust during this time (200% for 2011/12 and 2012/13), while *PLOS ONE*'s growth rate declined (in 2014 *PLOS ONE* experienced its first absolute decline in articles). The facts of this "case" are consistent with a story of competition between differentiated products. Given its "Nature" brand, and its price matching strategy, we would expect *Scientific Reports* to grow (at least partially at the expense of *PLOS ONE*).⁵²

However, more than 90% of our partners' publications still appear in subscription journals. If at some point (most) publishers successfully move from subscription Big Deals to OA Big Deals, the type of price competition implied by the above PLOS/Nature comparison will be compromised severely. Why? First, authors' journal publication choices will no longer have a pecuniary dimension (as is the case under the current subscription Big Deal regime). Second, in negotiating these OA Big Deals, institutions will face the same "multi-homing" problem as they do now: since collectively their authors demand the freedom to publish anywhere and have access to all content, most institutions can't credibly say no to either type of Big Deal contract.

Hence, in such an OA Big Deals world, the legal and economic arguments in support of antitrust activity would likely be far more persuasive. As mentioned previously, good theoretical and empirical support for a case is just a necessary condition for the initiation of an investigation. Active participation by the injured parties is also required to push things forward.

⁵² Davis, Phil. "PLOS ONE Shrinks by 11 Percent." The Scholarly Kitchen, January 7, 2016. <https://scholarlykitchen.sspnet.org/2016/01/06/plos-one-shrinks-by-11-percent/>

Appendix I: Brief Biosketches of Project Principals

Ivy Anderson, Director, California Digital Library, Oakland, CA

As Director of Collection Development and Management, Ivy Anderson oversees a broad range of shared collections activities on behalf of CDL and the ten UC campus libraries, with a goal of building world class shared collections available to all University of California students and faculty. CDL units under her leadership include Licensed Content (with staff based in Oakland and UC San Diego), which organizes and manages more than \$40 million in systemwide licensed resource expenditures annually; Mass Digitization, which coordinates large scale digitization partnerships with external partners such as Google and the Internet Archive as well as UC's participation in HathiTrust; and Shared Print, which facilitates the development of shared physical collections across the university and with extramural partners. Anderson oversaw the project's financial and bibliometric data collection and analysis, working with all of the library and other relevant partners and the modeling team.

Bo-Christer Björk, Hanken School of Economics, Helsinki, Finland

Dr. Björk is a Professor of Information Systems Science in the Department of Management and Organisation, Hanken School of Economics, Helsinki, Finland. From 1993 to 2000 he was professor of construction IT in the Royal Institute of Technology in Sweden, where he founded the *Electronic Journal of Information Technology in Construction*, an early OA journal. This led to an enduring research interest in the scientific research process which has been the focus of his work since 2000. Dr. Björk chaired the FinnOA committee from 2003-2008 and was a member of the board of the Open Access Scholarly Publishing Association (OASPA) from 2000-2012. He has an extensive open access publication portfolio includes numerous commissioned research reports for organizations including the British Library, the Max Planck Society Library, the Wellcome Trust, and others. With David Solomon (above), Dr. Björk recently co-authored *Developing an Effective Market for Open Access Article Processing Charges 2014*, a report commissioned by a consortium of research funders including Jisc, Research Libraries UK, Research Councils UK, the Wellcome Trust, the Austrian Science Fund, the Luxembourg National Research Fund and the Max Planck Institute for Gravitational Physics. Together with his colleague, Dr. Solomon, Dr. Björk advised the PIF project team on the financial and publishing data collection and analysis and served as a key contributor to the model design and development.

Mark McCabe, University of Michigan, Ann Arbor, MI

Dr. McCabe has appointments at the University of Michigan's School of Information and Boston University's School of Management. His current research interests include industrial organization, competition policy and regulation, and information economics. Dr. McCabe is an expert on the economics of journal publishing and has written several reports and articles on

the topic, including, “Online Access and the Scientific Journal Market: An Economist's Perspective,” a commissioned report for the National Academy of Sciences' Board on Science; and “A Portfolio Approach to Journal Pricing,” in the book *Economics and Usage of Digital Libraries: Biting the Bullet*. His work has been published in *American Economic Review*, *Nature*, *Rand Journal of Economics*, and *Journal of Academic Librarianship*, among other leading journals. He has received two Mellon Grants for his work on journal publishing – “Measuring the Impact of Digitization and Online Availability on Journal Citations,” (co-PI with Christopher Snyder, Dartmouth College), and “Scholarly Journals,” (co-PI with Daniel Rubinfeld, Aviv Nevo and Aaron Edlin, all at UC Berkeley). Dr. McCabe worked closely with both the qualitative and quantitative teams as an advisor on focus group and survey questions, identifying financial and publishing data for collection, and modeling techniques that incorporate both types of data.

MacKenzie Smith, University of California Davis Library, Davis, CA

MacKenzie Smith is the University Librarian at the University of California, Davis, charged with integrating digital resources and information technology necessary to support the academic community of the 21st Century. Smith is a long-time academic research librarian, specializing in technology and digital knowledge management. She previously worked for the libraries of Harvard and MIT, in Cambridge, Massachusetts, where she led cutting-edge projects on digital libraries and archives, such as the popular DSpace open source software platform; Web systems for online scholarly communication; and digital data curation in support of e-science. Smith has consulted widely in the library field, notably for the Association of Research Libraries to design and lead its E-Science Institute, and as a research fellow for Creative Commons to develop its strategy for sharing scientific research data and advocacy for open access to scholarship. Smith served as the principal investigator for this project, overseeing the project as a whole, the communication with the Foundation and the partnership, managing the budget (including subawards and contracts), participating in project teams, and conducting outreach to other libraries and related stakeholders.

David Solomon, Michigan State University, East Lansing, MI

Dr. Solomon is a Professor in the Department of Medicine and the Office Medical Education Research and Development at Michigan State University, with thirty years of experience in social science/educational research and evaluation. In 1996 he founded *Medical Education Online* (MEO), a respected peer reviewed web-based journal in medical education now published on an open access basis by Co-Action Publishing. Dr. Solomon is the author of *Developing Open Access Journals, A practical guide* (Chandos Publishing), and with other colleagues founded the Open Access Scholarly Publishing Association (OASPA), for which he served as a founding board member. Since 2011, Dr. Solomon has focused much of his scholarly work with Dr. Björk and others researching the nature and growth of open access publishing, particularly APC-funded OA publishing. Dr. Solomon worked closely with the project

team on the financial and publishing data collection and analysis, and was a key contributor to the model design and development.

Greg Tananbaum, ScholarNext Consulting, Kensington, CA

Greg Tananbaum is a scholarly communications consultant with nearly 20 years of experience at the intersection of technology, content, and academia. Other clients include Microsoft, Facebook, SPARC, the American Heart Association, Annual Reviews, and PLOS. Tananbaum served as the project's primary project manager and a member of the data modeling team, with additional responsibility for developing the publisher survey and the cost-per-article data.

Carol Tenopir, University of Tennessee, Knoxville, TN

Dr. Tenopir is a Chancellor's Professor at the School of Information Sciences at the University of Tennessee, Knoxville and the Director of Research for the College of Communication and Information, and Director of the Center for Information and Communication Studies. Her areas of teaching and research include: information access and retrieval, electronic publishing, and the information industry. She is the author of five books, including, *Communication Patterns of Engineers*, winner of the American Society for Engineering Education, Engineering Libraries Division 2005 Best Publication Award, (IEEE/Wiley InterScience, 2004) with Donald W. King. Dr. Tenopir led the project's qualitative data collection and analysis, in collaboration with the UC Davis team and members of the quantitative data team.

Matthew Willmott, California Digital Library, Oakland, CA

Mathew Willmott is the Scholarly Publishing Data Analyst at the California Digital Library, focusing primarily on this project since May 2015. Previously, Willmott worked in a multi-faceted role at the MIT Libraries, acting as the primary liaison between the library and the Physics Department, undertaking quantitative analyses of print and electronic collections across the library system, and developing and administering infrastructure supporting the implementation of the MIT Faculty Open Access Policy. Willmott was responsible for gathering, analyzing, and documenting data from library and bibliometric partners, and contributed to the analysis involved in developing the financial model.

Appendix J: Data Sharing Plan

Data collected and analyzed for the project has been deposited in the University of California's Dash data repository (<https://dash.cdlib.org/>) in the UC Office of the President's collection.

These data are made available under a Creative Commons license (CC-BY) unless otherwise noted in the data documentation.

- Author Survey Results
DOI: 10.5060/D8Z59F
URL: <http://n2t.net/ark:/b5060/d8z59f>
- Cost-Per-Article Analysis
DOI: 10.5060/D8G593
URL: <http://n2t.net/ark:/b5060/d8g593>
- Partner Institution Publication Volume
DOI: 10.5060/D86P4W
URL: <http://n2t.net/ark:/b5060/d86p4w>
- Partner Institution Research Expenditures
DOI: 10.5060/D8BC75
URL: <http://n2t.net/ark:/b5060/d8bc75>
- APC Payment and Pricing Data
DOI: 10.5060/D8301X
URL: <http://n2t.net/ark:/b5060/d8301x>
- APC - SNIP Regression Raw Data
DOI: 10.5060/D8KW2M
URL: <http://n2t.net/ark:/b5060/d8kw2m>