



UC DAVIS
UNIVERSITY OF CALIFORNIA



University of California
CDL
California Digital Library

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

University of California Libraries
September 19, 2014

Submitted by:

MacKenzie Smith, University Librarian
University of California, Davis
One Shields Avenue
Davis, CA 95616

Summary

The University of California, Davis, and the California Digital Library, on behalf of the University of California and its partners, seek a grant of \$800,000 to conduct an investigation of the economic implications of an 'Article Processing Charge' (APC) model for funding open access to scholarly journals. The proposed project would be conducted over fourteen months (twelve months for research and two months for write-up) during the eighteen month period from January, 2015 through June, 2016.

Table of Contents

- A. Background
 - B. Rationale
 - C. Project Description
 - D. Key Personnel
 - E. External Partners
 - F. Project Timeline
 - G. Outcomes, Benefits, and Intellectual Property
 - H. Reporting
 - I. Budget Narrative
 - J. Appendices
 - Bibliography
 - UC Board of Regents
- Supplementary Materials:
- CVs for Key Personnel
 - Partner Support letters

Background

In 2013, the Andrew W. Mellon Foundation granted the University of California, via the UC Davis campus and in collaboration with the California Digital Library, funding to develop a plan for an extensive and rigorous investigation of the economic implications of the “Article Processing Charge” (APC) funding model for open access to scholarly communication. 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)¹ 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. Our study focuses on APC-funded scholarship because we observe that should this model become the predominant open access funding model for the majority of journals in which authors at large North American research institution like the University of California publish, there is a significant risk that we would be unable to continue to sustain scholarly communication. We recognize that other forms of scholarship, e.g. monographs, are beginning to shift to a gold funding model as well, but journals are much further along in the transition and represent a much larger proportion of library budgets, so we are concentrating only on journals in this study.

We are also focused exclusively on large, North American research universities since one of the main motivations for this work is the observation that a relatively small number of U.S. research universities, among them the University of California, generate a substantial percentage of published research articles. There have been related studies done on this topic, but so far these have been focused on European research institutions that are primarily government funded and have much stronger government mandates to convert their publishing activities towards an open access model (Finch 2013).

The outcome of that 2013 planning effort is this proposal for a project to build a set of financial scenarios, or models, depicting the financial implications an APC-based system of scholarly journal publishing for the conversion of the current system of scholarly journal publishing to an APC-based system, for large research institutions. What follows is a detailed description of our proposed project.

Rationale

The key question that the proposed project asks is whether a large-scale conversion to open access scholarly journal publishing funded via APCs would be viable and financially sustainable for large North American research-intensive institutions, whose faculty currently author a significant percentage of the world’s research. For example, the University of California’s article share in the large publisher bundles licensed by the California Digital Library is typically in the vicinity of 2%; this percentage is often considerably higher in the most prestigious journals). We define ‘sustainable’ in this context as costing these institutions roughly no more than, and ideally considerably less than, current 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

¹ While we will examine the degree to which authors self-fund APCs, to the extent that institutions or grant funders, as proxies for authors, provide this support, it is our goal to understand the impact on current budgets of these costs – this is the primary motivation for this study. That said, author attitudes towards these fees have an impact on what publishers charge and that will be an aspect of the financial model we plan to build.

Project Proposal for *Pay It Forward: Investigating a Sustainable Model of Open Access Article Processing Charges for Large North American Research Institutions*

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. Our immediate interest is not in the cost of the scholarly publishing system as a whole – although we may be able to draw inferences about overall system costs – but rather the impact on representative large research institutions that, as key funders of the scholarly publishing marketplace, need to understand and help define any new financial model that is proposed to replace the current system.

Project Description

The project we have designed with this planning effort builds on work initially done by the California Digital Library to assess the implications on our budget of emerging APC levels and proposes an investigation to help us understand:

1. The overall cost to publish scholarly journals and the value of different components to editors and authors,
2. What the emerging APC model would cost UC and similar large research institutions under a variety of rigorously-modeled scenarios, and
3. Reasonable financial models for publishing research that cover basic costs and ongoing innovation.

The components of the project we envision are described below, including an overview of its goals and working methodology, the partnerships we have established to conduct the work, and our overall budget and timetable. The project we have planned will be conducted over 18 months, beginning in January, 2015 and concluding at the end of June, 2016.

Qualitative Analysis

The shift to an APC funding model implies the introduction of a new socio-technical system² for scholarly publishing and will produce changes in workflow for numerous stakeholders. 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 (Steiner 2008). 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 (de Waal and Batenburg 2014). In assessing stakeholder perceptions it is important to keep in mind that “value” is a multi-faceted and, often inter-subjective, concept that is directly linked to an individual’s unique social and institutional context (Bijker et al. 1997; Mol 2002; Pinch 2003). Given this complexity socio-technical design scholars suggest that sound research methodologies elucidate the perspectives of stakeholders, their actions and motivations, and relationships between stakeholders (Sjostrom and Goldkuhl 2009; Scholl 2004; Whitworth 2009).³

² 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).

³ Design research has been faulted for its emphasis on technological determinism without reference to social context resulting in high failure rates for newly introduced systems. Authors suggest that research should close the

Project Proposal for *Pay It Forward: Investigating a Sustainable Model of Open Access Article Processing Charges for Large North American Research Institutions*

Establishing a baseline understanding of stakeholders' perception of "value" and potential "buy-in" to a different scholarly publishing system is a necessary first step in our research plan. In order to understand potential barriers and facilitators in sufficient detail to accurately deploy these factors in our modeling, we will first need to identify the stakeholders⁴ and then assess their perception of the value derived from the current publishing paradigm versus the potential value possible in an APC model. Key stakeholders in this assessment include researchers (faculty and graduate students who both author and read publications), and publishers. While scholarly institutions, as represented by libraries in this project, are stakeholders in the scholarly communication system, their concern is financial sustainability⁵. This concern about financial sustainability will be addressed by the model(s) we build.

The data gathered from these key stakeholders will be used to inform and enhance the economic models we will be developing as the final product of this research. While a simple financial model of APC-funded journals is possible with just the financial and bibliometric data we will collect, in order to get at disciplinary differences in what costs journals can impose and where funds are available to pay APCs, we need to include parameters for author expectations. For example, while some fields are willing to publish in APC-funded 'mega-journals' regardless of their prestige, in other fields that is still unimaginable and authors will always select the highest impact journal regardless of cost. At this time there is a limited literature comprehensively addressing these matters with respect to a transition to APCs as this would specifically impact the North American research intensive university context. Moreover, these studies were conducted for other purposes and, while informative in designing this project, do not have the requisite detail and orientation required for our economic models⁶. Additionally, because the landscape of scholarly communication is changing so rapidly, the literature based upon data collected just a few years ago may not be appropriate to the current context and its use would negatively impact the accuracy of our models.

gap between social needs and technical performance or, in other words, what technology *does* and what communities *want* (Whitworth 2009).

⁴ See work by Gerald Steiner (2008) for a discussion of stakeholder identification and how stakeholder perspective can be employed in designing sustainable innovations. An article by Achterkamp and Vos (2007) focuses on stakeholder identification in the context of civil society.

⁵ For the purposes of this project we assume neutrality within the institutional setting and conduct analyses with regards to this stakeholder group's 'buy-in' at the level of the institution. Thus, the institution is considered a singular stakeholder entity whose primary concern is that any change in the scholarly publishing model does not result in increased, and preferably decreased, expenditures. This concern remains static regardless of how resources within the institution shift to pay APCs and does not address which unit should shoulder the costs (e.g., the Office of Research, academic departments, or libraries). These are valid questions, however, they are beyond the scope of this project.

⁶ One large-scale survey, the EU-funded SOAP Project (Dallmeier-Tiessen 2011) surveyed author attitudes about open access, including funding experiences and barriers. Another similar survey-based study was conducted in 2013 and drew from author-respondents publishing in Taylor & Francis Group and Routledge journals (Frass et al. 2013). Finally, work by Diane Harley and colleagues (2010) has investigated faculty attitudes towards the changing landscape of scholarly communication across seven different disciplines. However, the research by Harley and colleagues does not address the economic feasibility of various publishing models.

Project Proposal for *Pay It Forward: Investigating a Sustainable Model of Open Access Article Processing Charges for Large North American Research Institutions*

Assessing Researcher Perception of “Value” and Potential “Buy-in”

Data collection and analysis for the researcher stakeholders will be done by Professor Carol Tenopir (University of Tennessee), her research assistant, and Allison Fish (UC Davis postdoc) who will use qualitative methods, such as focus groups and surveys. Such a baseline will establish (1) How researchers value the publication of their own work; (2) How researchers value the accessibility of the scholarship that they generate; (3) How researchers value the production and accessibility of other scholars’ work; (4) How these attitudes vary amongst different disciplines; and (5) How these attitudes vary across different institutions. In particular, we want to assess how researchers currently understand their role in the current publishing paradigm (gated access), how they think this role might change in the event that the scholarly communication changes to an APC-funded model, and what perceived advantages and drawbacks might such a change produce. This will inform our understanding of the acceptability of APCs for this stakeholder group and may provide initial clues as to how advantages could be highlighted and drawbacks addressed in a successful and sustainable model.

In particular, we will gather qualitative data from researchers in different disciplines to assess what types and forms of scholarly communication are important to each field, whether APC-financed open access journals is a feasible goal for these different forms, and if there is widespread familiarity and experience with both open access publishing and APCs. For example, most health sciences researchers primarily produce article-length publications with multiple contributing authors and, given funding mandates of the NIH, are familiar with and have published in open access journals and databases (e.g., BioMed Central and PLOS). In contrast, researchers from the humanities and social sciences may have significantly less experience in this matter because fewer open access publishing platforms are available and there is less grant funding to draw on for support. In particular, many humanities disciplines consider book-length manuscripts to be the ideal vehicle for the dissemination of a fully developed scholarly argument, and few influential works have been successfully produced through open access mechanisms at this point. Therefore, there is more at stake for a humanities researcher since it would require him/her to publish the culmination of several years’ work in an untested manner.

A prior study on a similar topic has noted that author attitudes towards open access scholarly publishing is such a complex issue that it is difficult to conduct “meaningful research” in a single set of survey questions (Frass et al. 2013, 4). Therefore, the qualitative arm of our proposed project on APC-funded models will proceed in two phases. In the first phase we will gather data through focus groups with researchers from different disciplines at five sites across the four partner institutions. The sites where face-to-face focus groups take place are UC Davis, UC Irvine, The Ohio State University, the University of British Columbia, and Harvard University. These data will inform the second phase of the qualitative arm which will consist of anonymous surveys administered to the entire faculty of the five sites. It is common practice to use small, hand-selected focus groups to inform population-wide surveys when little understanding of a social phenomenon exists (Bernard 2006). Both the focus groups and surveys are described in greater detail below.

Human Subjects permission for both surveys and focus groups will be obtained from the University of Tennessee Institutional Review Board (IRB), with project partners gaining IRB permission as needed from

Project Proposal for *Pay It Forward: Investigating a Sustainable Model of Open Access Article Processing Charges for Large North American Research Institutions*

the institutions where research activities are taking place⁷. Specifically, in September 2014 Tenopir and her assistant will submit two sets of IRB applications at the University of Tennessee, one for the focus groups and one for the surveys⁸, in anticipation of this project. Tenopir expects that approval of the application will take approximately one month and will be obtained in October. Once approved, the University of Tennessee IRB documents will be forwarded to the project partners for inclusion in their own IRB submissions. We anticipate all participating institutions will be able to obtain IRB approval by January 2015 in advance of the focus groups, which will be initiated in the following month.

Rationale for Five Research Sites at the Four Partner Institutions

Both focus group and survey activities will take place at each of the four partner institutions. Three of these institutions (The Ohio State University, the University of British Columbia, and Harvard) have only one campus and, therefore, only one set of focus groups and one survey is required to obtain a representative sample of researcher attitudes at each of these locations. The fourth partner is the University of California, a system that has 10 geographically diffuse campuses. Therefore, a single set of focus groups and surveys will not be possible. For this reason, focus groups and surveys will be undertaken at two campuses: UC Irvine and UC Davis. These campuses were selected because they are of the same approximate size and ranking, have similar histories, and are noted as having researchers with active publishing portfolios. However, UC Irvine and UC Davis are also well matched in their differences. Specifically, UC Irvine is noted for its stature in the humanities, whereas UC Davis is noted for its stature in the sciences and engineering. Finally, UC Irvine is one of the early adopters of the UC Green Open Access policy, whereas UC Davis is in the second tier of adopters. For these reasons, UC Irvine and UC Davis, when combined, will provide a representative sample of UC faculty on the topics of open access and APC publishing models.

Focus Groups of Faculty Members and Graduate Students

Focus Group Purpose: Attitudes about open access and APC publishing models are complex and evolving, so focus groups will help provide a baseline for the contours and complexities of the issues. The focus group is a useful methodology when little is known about a particular topic or community and provides the researcher with a space for candid conversations not only about current practices and beliefs, but also about desirable or predicted changes into the future. Focus groups do not attempt to involve a representative sample of all viewpoints, but identify some current key behaviors and attitudes, and help bring out some strong beliefs and possible benefits or costs into the future. They also help suggest productive questions to be asked in follow-up surveys. These focus groups have two primary purposes: 1) to reveal current practices and opinions regarding open access and APCs within the focus group participants, to understand why they hold these opinions, and to determine what factors might change those practices or opinions, and 2) to inform the design of a survey sent to all researcher-stakeholders at the five partner sites. Focus groups are useful to get an in-depth look at motivations

⁷ We acknowledge the possibility that any of the partner universities may not receive IRB clearance to conduct the author focus groups or survey, and that in that case we will seek alternative institutions to insure an adequate pool of authors for this study.

⁸ We will request an exemption from the IRB for the anonymous surveys and anticipate that the focus groups will receive an expedited review process because there are no vulnerable populations involved.

Project Proposal for *Pay It Forward: Investigating a Sustainable Model of Open Access Article Processing Charges for Large North American Research Institutions*

and reasons for certain behaviors, while surveys will reveal how widespread opinions or behaviors are across a broader range of academics.

Focus Group Design: Two focus groups will be conducted at the five sites listed above; with one group faculty and one for graduate students⁹. The size of each focus group will be limited to approximately 10 participants¹⁰. Local contacts at each partner institution will be responsible for ensuring that each group is reflective of the campus and includes representation from multiple subject disciplines. This will result in a total of 10 focus groups with an average of 10 participants in each, for a total of approximately 100 participants. An incentive valued at approximately \$20-\$30 and refreshments will be given to compensate each participant for his/her time and participation.

Focus Group Data Collection: The objective of the focus group methodology is grounded in the idea that the conversation between participants is generative and allows for the community's concerns to emerge as opposed to the researcher designing a structured set of questions that assumes certain investments of the group studied. Therefore, focus groups will be conducted in a semi-structured manner with a limited set of questions and subject domains posed to elicit conversation amongst members of each group. Conversations will vary according to group, but will include domains and questions that are similar to those outlined below:

Domain 1: Focus group participants' access to other scholars' research

- How does access to high quality scholarly journals help you as an author, a researcher, a teacher, and/or a student?
- How do you currently get access to this high quality research?

Domain 2: Access to the products of focus group participants' research

- Why is it important that other researchers, teachers, and students have access to your publications?
- How do you think others get access to your publications?

Domain 3: Focus group participants' ideas on publishing costs and economic models

- What are the most common economic models that journals in your discipline use to sustain themselves financially? Are you aware of other models?
- What do you think are the best ways to pay for the costs of journal/article publishing? Who is the most logical party or parties to pay for these costs?

Domain 4: Focus group participants' opinions on open access and APCs

- Do you know what open access is and what are your thoughts about this model?
- Do you know what APCs are and what are your thoughts about them?

⁹ Focus groups will be separated by stakeholder type with one each for faculty and one for graduate students. This will be done to ensure that members of each group feel free to express their ideas. For example, graduate students might be reticent to talk in front of faculty members.

¹⁰ Ideally, focus groups will include no more than 12 participants and no fewer than 8 participants. Too many might inhibit some participants, whereas too few might fail to produce the necessary conversational atmosphere (Bernard 2011).

Project Proposal for *Pay It Forward: Investigating a Sustainable Model of Open Access Article Processing Charges for Large North American Research Institutions*

- Is there a difference in quality between open access/APCs and subscription-based journals? How is pricing related to quality?
- How might a change in pricing mechanism from subscription to open access help/hurt scholarship in your field, your institution, or for other researchers? Are there other viable alternatives?

Domain 5: Focus group participants' current publishing practices

- What types of venues do you publish in and how do you think this affects how others are able to access your research?
- Have you ever been asked to pay an author payment for an article at the time of publication? If so, what did you think? If not, would you be willing to?
- What are the most important criteria you use to determine where to publish and how do you see that changing in the future (if at all)?

Domain 6: Thoughts on the future of publishing

- Which methods of paying for journals would you like to support in the future?
- How can you/members of your stakeholder group influence cost models?

Carol Tenopir – an experienced moderator – will pose questions, keep conversations on track, elicit clarifications, and encourage contributions from each focus group participant. Each focus group will be audio-recorded and these recordings will be transcribed by the research assistants. Additionally, field notes of the focus groups will be kept and these can be used to inform the transcripts and catch unrecorded details of the discussions and research activities.

Focus Group Data Analysis: All three members of the research team will read each transcript and develop a code to identify common themes emerging from the focus groups. The research team will then meet, compare across their codes to develop a master code, and use NVivo to complete the analysis. Systematic analysis of focus group responses will help to develop and refine questions to be asked of a larger group in the survey. Commonly used words, ways of expressing terms, and sentiments regarding open access and APC concepts will be revealed in a comprehensive analysis of the discussions, as well as areas of confusion and concern.

Focus Group Reports: Initial insights from the focus group transcription and analysis will be used to inform the survey instrument design and refine the wording of the specific questions. A publishable report as a result of the focus group analysis will also form a distinct product and outcome from this project. The draft of this report will be shared first with the extended project team, with a final report produced thereafter and shared publically.

Surveys of Faculty and Graduate Students

Survey Purpose: Surveys are a good way to reach large numbers of participants to get candid responses to specific questions about current behaviors, attitudes, and opinions. They can be analyzed to reveal statistically significant differences among sub-groups that can be extrapolated to populations and subpopulations and ensure that a range of respondents are heard. The exact order and wording of questions on the survey instrument will be informed from the focus group transcripts and preliminary

Project Proposal for *Pay It Forward: Investigating a Sustainable Model of Open Access Article Processing Charges for Large North American Research Institutions*

analysis of focus groups. Confusion over definitions of terms, for example, and logical ordering of questions to match thought processes, are often first identified in focus groups.

Survey Design: Survey instruments will be developed and tested by Dr. Tenopir and other members of the project research team and distributed by partner institutions at the five research sites. Some of the questions will replicate questions that have been asked at other venues in the past (see, for example, Tenopir, et al., Nicholas, et al., Xia); other questions will be unique to this study. All questions will be designed to answer the key research questions of this project, including revealing familiarity, opinions, and current practices with open access and APCs and will be written with best survey practices in mind, including avoiding double negatives, leading questions, and illogical order (Miller & Salkind, 2002).

All questions will be pilot tested by approximately 15-20 researchers at the University of Tennessee and partner institutions to assure skip logic works correctly, wording is clear, that the flow of questions is logical, and to time how long it takes to complete the survey. To improve response rate, a drawing for an incentive will be offered to participants and the entire survey will be limited to no more than 15-20 minutes to complete.¹¹

Survey Data Collection: Project contacts at each of the participating institutions will contact their faculty and graduate students via email asking them to complete the questionnaire. An emailed invitation with an embedded link to the survey instrument (housed on the University of Tennessee Qualtrics server) will be sent out locally at each institution. All responses will be anonymous, but universities can choose to have a last question that will ask for an email address for those who wish to be entered into a drawing for an incentive. This identifying information will be stored separately from responses.

Survey response rate for academics has been decreasing over time (Beebe, et al. 2013), but we can anticipate an approximate 10-20% response rate for faculty members, with slightly lower rates for students. The approximate total population of faculty members in these five institutions is 19,000 and of graduate/professional students is 44,000. For a 95% confidence level with a confidence interval (margin of error) of +/-5% and a variance of .5, sample sizes of 377 faculty members and 381 graduate students are needed for the overall calculations (Babbie 2013; Smith 2014.) Further analysis will be broken down by broad discipline with up to five categories (e.g., humanities, sciences, social sciences, medical/health, and engineering/technology). Assuming an approximately equal number of faculty and graduate students across the broad disciplinary categories (or a subject population of approximately 3800 faculty members and 8800 graduate students in each category), we will need at least 100 responses per category to produce meaningful results at the 90% confidence level and 8% confidence interval. (Lower sub-population numbers require fewer responses per category.) If responses are not equally distributed across categories or numbers in any one category are low, then we will take several measures to allow meaningful analysis to be made. First, if response rates are disproportionately low in any subject areas, a follow-up message will be sent to faculty and graduate students. If response rates are still low and

¹¹ The evidence relating to response rate and length of surveys is inclusive, but a “lottery” style incentive is generally believed to be the most effective way to increase response rate. See, for example, Ziengenfuss, et al.)

Project Proposal for *Pay It Forward: Investigating a Sustainable Model of Open Access Article Processing Charges for Large North American Research Institutions*

respondents are not characteristic of the subject discipline populations, then results can be weighted proportionately (Babbie 2013) If population and respondent numbers are insufficient in any subject category, the subject discipline categories can also be redistributed—for example engineering/technology and medical/health can be analyzed together with sciences. The same is true for other independent variables that we may use in analysis, for example age or publishing productivity in 3-5 categories of relatively equal size with 100-200 respondents per category.

Since we will use the results of the focus groups to develop the surveys it is not possible to anticipate the exact questions used at this time. However, domains and types of questions that the survey will seek to answer include:

Domain 1: Demographics of survey participants

- Level of career (graduate student, researcher, faculty, other) and rank-in-level
- Disciplinary affiliation (text box)
- Age

Domain 2: Familiarity with open access and APCs

- On a scale of 1 to 10, please rate your familiarity with the concept of open access.
- On a scale of 1 to 10, please rate your familiarity with the concept of APCs.

Domain 3: Publishing practices

- How many articles did you publish in the last 5 years?
- What is the average number of co-authors for each of your articles?
- What percentage of the annual articles report on results of sponsored research, and do those sponsors consider publication costs allowable?
- Have any of these publications required you to pay any fees (e.g., page or color charges, submission fees, reprint fees)?
- How did you pay for these fees? (grant, research funds, personal, other)

Domain 4: Opinions on open access and APCs

- What do you think of the quality of articles published in open access venues?
- Does an APC indicate anything about the quality of the article?
- What do you think is the most appropriate source of open access fee funding (if any)?
- What range of costs/fees do you consider reasonable?

Survey Data Analysis: Survey results will be imported into SPSS for data analysis, which will be conducted by the project research team. Both descriptive and inferential statistical techniques will be used to answer the research questions. Frequencies and cross tabulations will reveal common behaviors and attitudes. These descriptive statistics, combined with standard tests of correlation to test for significance differences based on demographics such as subject discipline, age, productivity of publication, and familiarity with open access and APC publishing models, will provide an overall picture of publishing practices and opinions across research university faculty and graduate students. Research questions that will be answered with at least a 90% confidence level, based on number of respondents may include: are faculty members who publish more frequently less likely to publish in open access

Project Proposal for *Pay It Forward: Investigating a Sustainable Model of Open Access Article Processing Charges for Large North American Research Institutions*

journals (by segmenting the faculty respondents into low, medium, and high publishing)? Are faculty members in the sciences more favorably inclined towards author payments than faculty members in the humanities? Do younger faculty members have a lower threshold for a reasonable payment amount (by segmenting the respondents into 3 or 4 age groups)? Are PhD-level graduate students more likely to envision and articulate changes in publishing than professors?

Beyond the benefits described above, the inclusion of the behavioral and attitudinal data we will collect is important to our financial modeling for several reasons. It will allow us to understand the extent to which authors have the financial wherewithal to contemplate a growing APC publishing market, as well as the diversity of their funding sources. Further, by examining policies of major funding sources, we will be able to assess how flexible these funding bodies are in the application of grant monies toward APC charges and it will provide a snapshot of what authors are actually doing today with respect to APCs. The behavioral and attitudinal data to be gathered by the project will help inform the models developed in the final phase of the project. They will provide important context for the pricing levels and funding diversity within these models. This will extend the project beyond the realm of theoretical exercise and move it toward a set of prescriptions that have the chance of adoption within the authoring and publishing communities.

Assessing Publishers' Perception of "Value" and Potential "Buy-in"

Data collection and analysis for publisher stakeholders will be coordinated by key project personnel, Greg Tananbaum (ScholarNext) and Allison Fish (UC Davis postdoc), in collaboration with Association of Learned and Professional Society Publishers (ALPSP), a project partner. Survey methods will be used 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 want 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. This will inform our understanding of the acceptability of APCs for publisher stakeholders and may provide initial clues as to how advantages and challenges could be highlighted and addressed in a successful and sustainable model. In order to collect this information an anonymous survey methodology, discussed in detail below will be developed by Tananbaum, Fish, and ALPSP and disseminated amongst the latter's membership.

Rationale for the Selection of ALPSP as Project Partner: ALPSP was selected as a project partner because it is the largest international trade association of scholarly and professional publishers. ALPSP membership consists of 222 journal publishers from across disciplinary, geographic, and business model spectrums. In this endeavor, ALPSP will be represented by Suzanne Kavanagh (Director of Marketing and Membership Services, ALPSP), Mandy Hill (MD Academic Publishing, Cambridge University Press), and Audrey McCulloch (Chief Executive, ALPSP).

Survey Purpose: As stated above, in the section describing surveys for researcher stakeholders, anonymous surveys are a good way to reach large numbers of participants and can be analyzed to reveal statistically significant results that can be further extrapolated to larger populations of which little is

Project Proposal for *Pay It Forward: Investigating a Sustainable Model of Open Access Article Processing Charges for Large North American Research Institutions*

known. The exact order and wording of questions for the survey sent to publisher stakeholders will be developed by the ALPSP representatives and Tananbaum and geared towards developing baseline understanding of the issues delineated above.

Survey Design and Data Collection: An attitudinal and behavioral survey will be designed and tested by ALPSP, Tananbaum, and Fish to produce better information of how the growth in APC publishing practices has impacted publishers' operations, their current business models, and their organizational planning. Once the survey design has been finalized ALPSP will oversee its execution (via SurveyMonkey or a comparable online survey tool) and send out participation requests to all ALPSP members with active journal publishing programs. The survey will be sent under ALPSP's imprimatur, which will encourage high levels of participation among its members, and our publishing partners (Elsevier and Thomson Reuters) will also encourage participation. In order to ensure candid feedback ALPSP will anonymize individual responses. This, combined with the fact that the survey will be run by ALPSP, should assuage any concerns about the sharing of potentially sensitive information. The parties anticipate a response rate in excess of 10% of publishers, to include larger publishers that account for as much as 50% of all published research.

The survey will seek publisher input within the primary domains and sample queries are listed below.

Domain 1: Characteristics of publisher.

- Organization type (not-for-profit, for profit)
- Approximate size (number of employees and journal roster)
- Location

Domain 2: Current cost model used by publisher.

- What cost model do you current employ (% of journal roster that is subscription, APC, or hybrid)?
- If you do offer APC options, do you charge per article, per page, or per author? How much do you charge?
- If you offer APC options, do you also offer APC waivers to authors who require them? How extensive is the need to offer waivers, and how do you factor this into the fees that you charge and your growth projections over time?

Domain 3: How current trends in APCs have impacted current operations.

- Are the tasks that publishers perform changing noticeably under this emerging model?
- Are resources being reallocated at nontrivial levels from certain departments/tasks to others?

Domain 4: How current trends in APCs are expected to impact future operations.

- To what extent, if any, have publishers planned for the continued growth in the APC model?
- What level of organizational reorientation is this change likely to require?

Domain 5: How publishers feel about the APC model.

- How is the evolution of the traditional journal publishing business model seen as an opportunity, and how is it perceived as a threat?

- What is the appetite within the publishing community for exploring alternative models?

Survey Data Analysis: ALPSP will, as stated above, strip identifiers before forwarding raw results to Tananbaum and Fish who will conduct statistical analyses to discern trends and themes. To the extent that the response pool allows, we will draw meaningful conclusions by publisher type (e.g., disciplinary focus, location of journal). In its totality, the ALPSP survey will provide crucial insight into (a) how publishers perceive “value” within both the current and proposed APC-centric business models; and (b) the likelihood that publishers will “buy-in” to an APC-centric model without significant disruption to the traditional model of research dissemination.

Assessing Institutional Perception of “Value” and Potential “Buy-in”

Scholarly institutions are important stakeholders in the scholarly communication and publishing system and, in the project, are represented by the library contacts of the partner institutions. However, as stated above, for the purposes of this research, the primary concern governing institutional perceptions of “value” and potential for “buy-in” is whether or not a specific model is deemed to be financially sustainable from the university perspective (i.e., does it result in increased, decreased, or equivalent costs to the university as a whole). This concern relating to financial sustainability, as a cost to the university as a whole, is the crux of the model(s) we will build as an end product of this research.

Financial and Bibliometric Analysis

Several prior analyses have suggested that a publishing system relying on APC-funded publications would not be sustainable for large research institutions, given their large share of the overall research output, and this has led many observers to discount the viability of broad-based adoption of the APC funding model for journals (see for example Davis et al 2004; Báscones Domínguez 2006; Walters 2007). However, these analyses have often relied on simple calculations that fail to take into account the multiplicity of factors at play in an APC-funded publishing model, such as co-authorship patterns and the availability of sponsored research funding. Nor have they taken into account authors’ and institutions’ attitudes and beliefs about APCs, in general and among different disciplines. Recent studies commissioned in the U.K., where a transition to Gold Open Access is a stated (if controversial) policy objective, conversely suggest that an APC-funded publication model would be less expensive for research institutions if these elements are considered (e.g. Swan 2010; Swan and Houghton 2012). Our goal is to expand and stress-test these assumptions by building a rich model that takes into account at least the following factors:

1. The current state of APCs as they are evolving today, including recent trend data,
2. The attitudes and behaviors of authors in various disciplines, including their willingness to publish under an APC model,
3. Changes in the global publishing environment that might reasonably be expected to lower the cost of publication, such as new technologies enabling lower overhead, elimination of cost categories such as print publication and access management infrastructure, and the likely range of APCs that might result from such changes,
4. The role of research funders in contributing to the cost of publication
5. Levels of co-authorship in many disciplines, which are likely to spread the cost of publication across multiple institutions,

Project Proposal for *Pay It Forward: Investigating a Sustainable Model of Open Access Article Processing Charges for Large North American Research Institutions*

6. The introduction of healthy market forces that can moderate pricing, and
7. Disciplinary differences in all of the above areas that must be factored into any well-developed model.

To understand the relationship of current institutional expenditures to potential costs under an APC model, the data analysis and modeling phase of this project will assemble comprehensive data about current subscription expenditures at a representative set of major North American research-intensive institutions, and will marry that with data about the journal publication output at those same institutions and in the world at large. This will be accomplished in four overlapping phases of work, described below:

1. Gathering Institutional Expenditure Data
2. Gathering Publication Data
3. Gathering Current APC Data
4. Scenario Development and Modeling

Institutional Expenditure Data

For this phase, we will collect comprehensive annual subscription and licensing expenditures for scholarly research journals from each library partner for the five-year period from 2009 to 2013. A rigorous methodology will be employed to eliminate costs likely to remain unaffected by a conversion to APC-funded publication, such as archival subscriptions and secondary full-text aggregations. Five years of data will allow us to perform trend analysis, i.e. to compare the results of the APC modeling to overall journal expenditures and to the growth rate of those expenditures over time. At the end of the project we will be able to answer the question “How much more or less expensive is a given APC funding model likely to be overall, both now and over time, compared with what research institutions are spending as whole for scholarly journals now?”

Each partner institution is asked to provide information about two categories of journal expenditure:

1. Total serials expenditure, both print and electronic, in each of the 5 years, excluding from the total expenditures for non- journal content and for some journal content that might be presumed to continue on a subscription basis under an APC model
2. Expenditures for open access publication that may exist outside of the traditional serials budget (e.g. open access memberships and campus publication funds)

Serials expenditure data will be collected according to the following methodology:

Exclusions: The following categories of serial expenditure will be excluded from the expenditure totals used for modeling purposes:

- Databases, e-books, and other non-journal content that may be funded via the serials budget
- Secondary aggregator databases (EBSCO, ProQuest, Gale, Factiva, Lexis-Nexis, etc.)
- JSTOR retrospective collections and other archival purchases or subscriptions
- Newspapers and other news sources

Project Proposal for *Pay It Forward: Investigating a Sustainable Model of Open Access Article Processing Charges for Large North American Research Institutions*

- Popular and trade journals and magazines
- Other content categories that may be identified

If possible, expenditures for each year will be broken out by print vs. electronic in order to make it possible to analyze the impact of retaining or discontinuing print in the subsequent models. Combined print and electronic subscriptions will be reported separately or counted as electronic, to be determined in consultation with the partners. If these expenditures cannot be broken out, we will determine how to treat this variable in consultation with the library partners, e.g., ask for general estimates of the percentage of print journal retention, determine the likelihood that each institution would choose to retain or forego print, and factor this into our models as a potential range of variation.

Expenditures for a small number of known journal packages (e.g. ACM Digital Library, IEEE Xplore Digital Library) may need to be adjusted to account for non-journal content in those packages that would continue to be licensed on a subscription basis if the journal content were converted to open access. To control for this factor, the publishers will be asked (under the auspices of our partnership with ALPSP) to provide the percentage of cost attributable to the journals in those packages; each library partner will be asked to calculate the cost of the journals in its package using that percentage.

Analyzing journal expenditures by discipline:

The institutional partners would like to understand not just the potential impact of APCs on their bottom line journal expenditures, but also how expenditures by discipline might change under an APC-funded model and how this aligns with disciplinary culture and funding patterns. However, tying journal expenditures to disciplines is not straightforward due in large part to the bulk pricing structure of large publisher packages. We will work with a subset of one or more partners on an experimental basis to analyze 2013 expenditure data by discipline, and incorporate this into our findings if the data appear reliable. Our preliminary methodology for this, subject to partner review, will be to ask the sample partners for detailed expenditure data by journal where these data are available (i.e. for single subscriptions or packages with line item detail), allocating the subscription cost to the subject area covered by the journal (using the same Scopus disciplinary taxonomy that will be used for publication output data below). Where only package expenditure data is available, we will analyze the package holdings by title, determine from that the disciplinary distribution of the journals in the package, and using current publisher list prices, allocate a corresponding percentage of the package cost to each discipline. This is a methodology that was successfully used to calculate the redirection of subscription costs for journals participating in the SCOAP3 high energy physics open access initiative. However, there are several potential pitfalls in this methodology when applied to a large set of holdings over time which lead us to consider this experimental: for example, it presumes that we can reliably identify the journals in each package held by the library partner, that we can obtain corresponding list price data from what is likely to be a large set of affected publishers, and that we can map that data to partner holdings. Since journal holdings change over time both within libraries and among publishers, we will perform this analysis for only the latest one-to-two years at best and assume that the disciplinary expenditure distribution thus derived is reasonably constant over the five year period, however this

Project Proposal for *Pay It Forward: Investigating a Sustainable Model of Open Access Article Processing Charges for Large North American Research Institutions*

assumption may warrant testing. Since some partner expenditure data will be covered by confidentiality agreements, we will explore this methodology first with the UC partners only, where access to detailed expenditure data at the journal and package level is not legally restricted. Other partners will be invited to execute the same strategy if they have the resources to do so, however we will not require this due to the complexities involved.

Publication Output and Authorship Data

The publication data used to construct our models will be obtained via a joint partnership with both Scopus and Web of Science. Scopus is envisioned as the project's primary data source because it indexes a larger number of scholarly journals than Web of Science (some 20,000 titles compared to a little over 12,000 titles in Web of Science), including more journals in the humanities and social sciences.

Journal Coverage by Subject, approximate as of January 2014¹²

	Arts & Humanities	Social Sciences	Sciences
Web of Science	1,725	3,114	8,613
Scopus	2,621	5,775	23,668

Web of Science, by contrast, contains more comprehensive sponsored research information, particularly in the sciences; Web of Science funder acknowledgement data goes back at least to 2009, whereas this data is only reliably captured in Scopus from 2013 forward. Together, these two data sources will allow us to assemble the most comprehensive picture of research output possible. We will account for differences in the two data sources via appropriate crosswalks (e.g. between different disciplinary taxonomies) and other adjustments that may be needed. The active participation of both data providers will ensure that differences are identified and addressed to the extent possible.

Two sets of publication data will be acquired for modeling purposes, principally from Scopus: general data about publication output, by discipline; and publication data specific to each institutional partner. Broad-based publication data will allow us to discern disciplinary differences in publication practices such as publication volume and patterns of co-authorship and will serve as a baseline to determine whether the publication characteristics of the partner institutions are broadly representative or exhibit unusual traits. It will also allow us to make inferences about the percentage of worldwide production

¹² Source: University of California Libraries, *Scopus Task Force Report* (September 2014 - internal document). Includes both active and inactive titles. Per personal communication from Scopus: Scopus now contains 2,527 active journals in Arts & Humanities, of which 1,509 also covered in Web of Science. For more information about the development of arts and humanities content in Scopus, see Meester, Wim, "Towards a comprehensive citation index for the Arts & Humanities," *Research Trends*, no. 32 (March 2013). <http://www.researchtrends.com/issue-32-march-2013/towards-a-comprehensive-citation-index-for-the-arts-humanities/>. Additional information about content selection in Scopus can be found in the Scopus Content Coverage Guide at http://www.elsevier.com/__data/assets/pdf_file/0019/148402/SC_Content-Coverage-Guide_July-2014.PDF

Project Proposal for *Pay It Forward: Investigating a Sustainable Model of Open Access Article Processing Charges for Large North American Research Institutions*

accounted for by large North American research intensive institutions and, thus, the financial burden such institutions might assume in an APC-funded system. The specific data to be acquired include the following (generally from Scopus, unless indicated otherwise):

1. *General Publication Data*: By Scopus discipline, by year of publication 2009-2013, for all research journal and proceedings content – excluding trade publications, books, etc.:
 - a. # of research articles
 - b. # of non-research articles
 - c. Histogram of the distribution of research articles vs. other content items, as defined in Scopus’s document typology (e.g., excluding editorials, errata, letters, reviews, etc.) across all journals in that discipline, to discover whether significant non-research-article content tends to be concentrated in a small number of journals within a discipline, or is broadly characteristic of the discipline. Since non-research-article content must be supported in any business model, understanding the distribution of this content will be important in suggesting how that content might be funded in an APC-funded system.¹³
 - d. # of research articles that are sole-authored vs. multi-authored
 - e. # of research articles that have co-authors from multiple institutions
 - f. Average and median # of co-authors per research article
 - g. # of research articles that are the result of sponsored research (from Web of Science)
 - h. # of open access journals
 - i. # of open access articles (sampling techniques will be required in order to estimate the number of OA articles in so-called ‘hybrid’ journals)
2. *Institutional Partner Data*. For each institutional partner, by discipline, by year of publication 2009-2013:
 - a. # of research articles with at least one author from that institution. From that set:
 - b. # of research articles that are sole-authored vs. multi-authored
 - c. # of research articles that have co-authors at multiple institutions
 - d. Average and median # of co-authors per research article
 - e. # of research articles that are the result of sponsored research (from Web of Science)
 - f. # of open access articles (sampling techniques will be required in order to estimate the number of OA articles in so-called ‘hybrid’ journals)

¹³ Scopus attempts to map the document type classification of the publisher to its own classification scheme. For example, “Letters” in a journal such as *Nature*, which are actually short research articles, are assigned the document type “Article” in Scopus. Scopus defines articles in its document typology as follows: “Original research, or opinion, also includes conference papers. Characteristics: in peer-reviewed journals, articles are usually several pages in length, most often subdivided into sections: abstract, introduction, materials & methods, results, conclusions, discussion and references. However, case reports, technical and research notes and short communications are also considered to be articles and may be as little as one page in length.” See Scopus Content Coverage Guide, p. 10 at http://www.elsevier.com/__data/assets/pdf_file/0019/148402/SC_Content-Coverage-Guide_July-2014.PDF. The analysis team will review both the Scopus and Web of Science document typologies more fully to determine whether the histogram approach or some other method will yield the best data about the distribution of research vs. non-research content.

- g. # of research articles whose Corresponding Author is from the partner institution vs. another institution
 - h. For the most recent 1-2 years, a list of journals in which authors at the institutional partners have published, with unique article counts
3. *Medline Data*. Twenty percent (20%) of Medline titles indexed by Scopus include the article's first author only. This is a potentially important gap, since co-authorship patterns are a key element of our study. We will seek to acquire missing Medline co-authorship data from Web of Science if available; if not available, we will exclude the affected titles from our sample for modeling purposes. We are still assembling information about the affected titles and will determine a course of action based on that analysis.

Gathering APC Data

An analysis of current article processing charges as they are developing in the marketplace will serve as an important baseline for our modeling. Typical publication fees for hybrid OA journals – that is, subscription journals in which selected articles can be published as open access with payment of an APC) are commonly estimated at \$3,000 for example; while these fees have often been used as a benchmark for APCs in general, some of the largest publishers are now differentiating these fees by journal and by discipline. Newer entrants into the open access publishing sphere with lower cost structures and thus lower publication fees are also radically altering the APC environment. Any modeling needs to take these current trends into consideration.¹⁴ Prior studies by Solomon and Bjork (Solomon and Bjork 2012; Bjork and Solomon 2014) extensively researched APC costs, but most of this information dates from 2011. The current project proposes to update this work to 2014 for the journals in which authors at the library partner institutions publish. This will not only provide a more current baseline keyed to the publishing behaviors of the large research institutions under study, but will also provide information about trends in APCs as they are developing in the laboratory of the marketplace. Solomon and Bjork will build on their prior work for this project and will correlate the data with Scopus disciplinary categories and impact measures to produce a map of likely APCs for the institutional partners differentiated by discipline and impact.

Scenario Development and Modeling

Working in consultation with a research economist (Mark McCabe) and under the direction of the core project team (Tananbaum, Anderson, Smith, and Farley), David Solomon and Bo-Christer Bjork will serve as modeling consultants for the project. Solomon and Bjork will incorporate all of the data inputs developed in this project – institutional expenditures, Scopus publication data, current APCs, potential future APCs derived from our research on publication costs, and information on author needs and preferences gleaned from author surveys and focus groups – to build a series of models depicting the

¹⁴ Journals that charge APCs typically offer waivers to individuals who are unable to pay the article fee. These subsidies are factored into the APC that a given publisher charges at the time the fee is established; therefore our study will consider the impact of fee waivers in the phases of work devoted to surveying publishers and analyzing publishing costs. Because authors at the institutions that are the subject of this study are unlikely to require such waivers themselves, we do not anticipate a need to address waivers directly in the modeling of institutional costs unless this emerges as a consideration in the author surveys. We will, however, include questions about waiver practices in the ALPSP survey of publishers.

Project Proposal for *Pay It Forward: Investigating a Sustainable Model of Open Access Article Processing Charges for Large North American Research Institutions*

social and financial impact of a largely APC-funded publishing landscape on the institutions participating in the study under a variety of assumptions including levels of grant funding, co-authorship patterns, realistically envisioned APCs that take into account the analysis of publication costs, and author uptake of APC publishing venues in different disciplines.

Financial Modeling

As previously mentioned, a key goal of the project is to construct a rich model that takes into account the current research output at large North American institutions, the ways in which a transition to a fully APC-based model might impact the journal publishing expenditures of these institutions, and how sustainable and desirable this fully APC-based model might prove for not only the institutions, but also for authors and for publishers.

A key component of this model construction is therefore the enumeration and costing out of key tasks associated with publishing scholarly journals. This is necessary to determine whether an APC-based model that is sustainable for large North American institutions and for authors is also viable for publishers.

We intend to develop a ground-up cost model for publishing in different disciplines. This modelling will draw from the experiences of scholars and departments from the partner institutions, including UC Davis neuroscientist Michael Rogawski, who is founding a new journal of negative results called *Epilepsy Reports* with the Society for Neuroscience, and the California Digital Library's eScholarship unit, which manages a collection of 65 electronic journals. Other sources of this type of information will be consulted as appropriate, such as the directors of PeerJ, Ubiquity Press, and other new online publishers that have managed to dramatically lower costs for APC-based publishing. The model will also rely on available literature (e.g., 73 Things Publishers Do; "Launching (and Sustaining) a Scholarly Journal on the Internet: The International Journal of Baudrillard Studies", and Waltham 2004), and the project manager's direct experience in launching dozens of scholarly journals across multiple disciplines.

The goal of the ground-up cost model is to accurately reflect the full range of expenses that are necessary to run an APC funded journal. These expenses include salaries and benefits, editorial, technical, operations, sales and marketing, and administrative. The model must also address the notion of profits, and how these profits can be used to fuel growth and innovation, as well as related costs such as long-term preservation (Waltham 2004).

We chose this approach for several reasons. We explored the possibility of acquiring actual line item expenditure data from both individual publishers and through publishing trade organizations, but were unable to do so. Publishers are reluctant to share this information on the grounds that it is proprietary and might be misrepresented or misunderstood. In addition, discussions with a number of publishers revealed that the quality of the cost data they maintain is inconsistent and varies widely across publishers. These inconsistencies would make it difficult to draw accurate conclusions based on such data even if it could be readily obtained. Finally, existing publishers have 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. As such, any modelling based on existing costs would encounter a number of legacy

Project Proposal for *Pay It Forward: Investigating a Sustainable Model of Open Access Article Processing Charges for Large North American Research Institutions*

expenditures that would be difficult to weed out from essential operations. A ground-up model provides a more accurate window into what the true costs of a fully APC-based model would be. We will also compare components of this model with author attitudes and preferences about which publishing services they value.

Key Personnel (CVs are provided with the Supplementary Material)

MacKenzie Smith, University Librarian, UC Davis University Library

MacKenzie is a senior library leader with a background in technology and new forms of scholarly communication. She will serve as the principal investigator for the 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.

Laine Farley, Executive Director, UC California Digital Library

Laine will be the project's co-principal investigator, overseeing the work at the California Digital Library and the library partner relationships for data collection and analysis.

Ivy Anderson, Director of Collections, UC California Digital Library

Ivy will oversee the project's financial and bibliometric data collection and analysis, working with all of the library and other relevant partners and the modeling team.

Scholarly Publishing Data Analyst (TBH), UC California Digital Library

A Scholarly Publishing Data Analyst will be hired by the CDL to work for Ivy Anderson on the project's data collection, cleanup, normalization, and analysis tasks. The data analyst will have primary responsibility for collecting journal subscription and OA expenditure data from the library partners and for coordinating the work of the quantitative analysis partners to facilitate the integration of expenditure and bibliometric data into the financial models. Responsibilities will include developing templates and tools needed to support data collection and reporting, performing expenditure data cleanup and normalization, acquiring relevant pricing data from publishers as needed, investigating and working with the partners to resolve data issues, and documenting data analysis methodologies for dissemination purposes. The full position description is included as an Appendix.

Allison Fish, Scholarly Communication Postdoc, UC Davis University Library

Allison is an anthropologist and JD, working with the UC Davis University Library on the transition of scholarly communication to open access models and the consequent behavioral changes. She will coordinate the project's qualitative research work with Professors Tenopir and McCabe and, in particular, organize the University of California involvement in faculty and graduate student focus groups and surveys. She will also assist Tenopir in the design, analysis, and interpretation of focus group and survey results for inclusion in our financial model. Additionally, she will coordinate with Greg Tananbaum and ALPSP on the publisher survey (e.g., drafting survey questions and processing results). Finally, she will conduct user feedback sessions with the library directors and other potential users of the model that this project produces and assist the write-up of results for publication.

Project Proposal for *Pay It Forward: Investigating a Sustainable Model of Open Access Article Processing Charges for Large North American Research Institutions*

Esther Hernandez, Grant Administrator, UC Davis University Library

Esther is an administrative specialist with the UC Davis University Library who will be handling the project contracts and subawards, budget oversight, personnel and reporting requirements (other than the Foundation reports). She will also coordinate overall administrative support for the project, such as scheduling conference calls, processing travel reimbursements, organizing in-person meetings, etc.

Greg Tananbaum, Consultant, ScholarNext

Greg is a scholarly communications consultant with nearly 20 years of experience at the intersection of technology, content, and academia. Greg will serve as the primary project manager and a member of the data modeling team, with responsibility for developing the ground-up publishing cost model.

External Partners

In addition to the key personnel listed above (the same team who performed the earlier planning effort) we have identified an extensive set of partners necessary to conduct the work we envision for the proposed project. These span the key stakeholder communities and expertise required to conduct the research we described and to build the models that will be the main outcome of the project. All of the partners listed below have agreed to participate in the project, if funded, and have developed Statements of Work and budgets to support their participation. A summary of the role of each partner is included below.

University Partners

While the project is based in the University of California and will involve its ten campuses and the CDL, we feel it necessary and useful to include additional research institutions for several reasons: to validate the data that we produce for UC, to compare our faculty's attitudes and behaviors to others, and to bring additional and complementary library and institutional expertise to the project. 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). After due consideration of the many institutions that we could consider inviting as partners, we chose three:

Harvard University Library, Cambridge, MA, USA (led by Lauren Syer, Manager, E-Resources Unit)

Ohio State University Libraries, Columbus, OH, USA (led by Karla Strieb, Associate University Librarian for Scholarly Communication and Tech Services)

University of British Columbia Library, Vancouver, BC, Canada (led by Jo Anne Newyear Ramirez, Associate University Librarian, Collection Development and Management Programs)

Qualitative Analysis Partners

Carol Tenopir, University of Tennessee, Knoxville, TN, USA

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.

Project Proposal for *Pay It Forward: Investigating a Sustainable Model of Open Access Article Processing Charges for Large North American Research Institutions*

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 will lead the project's qualitative data collection and analysis, in collaboration with the UC Davis team and members of the quantitative data team (Bjork, McCabe, and Anderson).

Association of Professional and Learned Society Publishers (ALPSP, led by Audrey McCulloch, Executive Director

ALPSP is an international association representing all types of nonprofit publishers and is the largest trade association for scholarly and professional publishers with more than 300 members in 40 countries, publishing scholarly content in many different ways. ALPSP membership includes a wide range of different types of publishers - journal publishers, book publishers, learned societies and professional bodies, database publishers, university presses and intergovernmental organizations. The variety and range of publisher members and associate members provides a unique network serving the interests of scholarly publishing worldwide.

Quantitative Analysis Partners

Scopus (Elsevier)

As a key source of data and analytics expertise, we will partner with Elsevier's Global Academic Relations (GAR) team, a non-commercial arm of the company that has extensive access to and knowledge of the Scopus database on which much of our research will depend. The University of California subscribes to Scopus and has access to this data directly, but we feel that the partnership with the GAR team will provide critical data analysis expertise that we can independently verify with our other partners. The GAR group has committed to partnering with UC Davis/CDL on this project as an unfunded partner, with Brad Fenwick as the lead coordinator.

Web of Science (Thomson-Reuters)

Thomson-Reuters' Scholarly & Scientific Research unit, which is responsible for the Web of Science database, is another key source of data and analytics expertise. We will partner with the Thomson-Reuters Web of Science team to provide data unavailable from Scopus, including funder acknowledgment data from 2009 forward. Partnering with Web of Science will also provide another locus of analysis and verification to ensure the reliability of our data. Thomson-Reuters has committed to partnering with UC Davis/CDL on this project as an unfunded partner, with Chris Burghardt as the lead coordinator.

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

Project Proposal for *Pay It Forward: Investigating a Sustainable Model of Open Access Article Processing Charges for Large North American Research Institutions*

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 will work closely with CDL on the financial and publishing data collection and analysis, and be a key contributor to the model design and development.

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. He possesses advanced degrees in systems science, economics, and construction management. 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 will advise CDL on the financial and publishing data collection and analysis and be 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: Byting the Bullet*. Dr. McCabe has published numerous articles on various component of journal publishing, including open access, online availability and citation counts, and journal pricing. 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 has also received grants on this topic from the Sloan Foundation, the Open Society Institute, and the American Association of Law Libraries. Dr. McCabe will be working 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.

Project Proposal for *Pay It Forward: Investigating a Sustainable Model of Open Access Article Processing Charges for Large North American Research Institutions*

Project Timeline (Leads for each component are indicated in parentheses)

While we envision the project conducting the majority of our work during the calendar year 2015 and spending another two months (i.e., by March, 2016) to summarize our results, given inevitable delays due to vacations and unforeseen absences, having the grace of an additional 6 months minimizes the risk of needing an extension.

January-March 2015

- Execute hiring agreement with external consultants and issue subawards to partner institutions (Smith, Farley);
- Establish project budget and tracking mechanism (Smith);
- Establish team communication processes: internal weekly, extended bi-weekly (Tananbaum);
- Hire project staff (Anderson);
- Review/finalize data collection methodology (Anderson, Tananbaum, McCabe);
- Collect data from CDL sources (Anderson);
- Acquire initial data from Scopus and Web of Science (Anderson);
- Define data collection procedure for UC campuses, UBC, OSU, and Harvard (Anderson);
- Complete IRB process at UTK, UC campuses, and partner institutions (Tenopir, Fish);
- Conduct focus groups at UCD, UCI, UBC, OSU, and Harvard (Tenopir, Fish);
- Conduct literature review on current APC pricing and trends (Solomon)
- Collect and collate APC pricing and expenditure data from existing sources, by gathering updated data to address gaps (Solomon)

April-June 2015

- Finalize and conduct faculty/grad student survey at UCD and UCI, UBC, OSU, and Harvard (Tenopir, Fish, McCabe);
- Conduct publisher survey via ALPSP (Tananbaum, Tenopir, McCabe);
- Complete data collection from UC campuses, UBC, OSU, and Harvard (Anderson);
- Refine data collection from Scopus and Web of Science, as needed (Anderson)
- Complete gathering of APC pricing and expenditure data (Solomon)

July-December 2015

- Complete analysis of survey data (Tenopir, Fish, McCabe);
- Complete analysis of financial data (Anderson, Tananbaum);
- Build version 1 model (Solomon, all);
- Review model with extended team (Anderson, Tananbaum);
- Finalize model (Solomon, all);

January-February 2016

- Review/refine model with UC and partner library colleagues, other stakeholders (Anderson, Fish, all);
- Write and publish final report (Smith, all)

March-June 30 2016

- No planned activities unless the project timeline has changed and with the Foundation's prior approval

Outcomes and Benefits

The main outcome of the proposed project will be a detailed, flexible, and publicly accessible financial model (or set of financial scenarios). The model(s) will be described in detail in a report to be written by the project team, shared publically, and disseminated via multiple published articles, authored by the key personnel and our expert consultants and collaborators, and via conference presentations at ARL and elsewhere. The project personnel are all frequently invited speakers in the library, higher education, and open access communities and we have presentations lined up for key events (e.g. the OAI conference in Geneva) if we are funded to begin the work next year.

We may additionally create spreadsheets for the model that can also be publically shared and disseminated, to save organizations the time necessary to operationalize the model(s) themselves. The advisability for that will become clearer as the parameters for the model emerge and we understand what form it should take. The UC Davis postdoc (Fish) will also be studying the utility of the model(s) for library directors and staff in the project's partnership, so that we can insure its utility before we conclude our work.

As for benefits, beyond the model as a new tool for university administrators and librarians to use in developing open access policies and strategies, there are significant potential costs associated with *not* having the information provided by this study available and, as a result, having to learn incrementally and from perhaps costly mistakes. In short, a relatively small investment now is likely to save a great deal later.

Intellectual Property

The main outcome of the proposed project will be a detailed, flexible, and publicly accessible financial model or set of financial scenarios that will be published online under an appropriate CC or similar license (e.g., data analysis software tools under a BSD license, reports under a CC-BY license, including descriptions of the methodologies developed for the project, and aggregated/analyzed data under a CC0 waiver). The project will rely extensively on data sourced from Scopus (Elsevier, B.V.), and Web of Science (Thomson-Reuters) that are the exclusive intellectual property of their companies and cannot be shared publicly, nor will we share detailed financial data from our partner institutions that fall under non-disclosure licensing terms. However we will document the data analysis methodologies requested of Elsevier and/or Thomson-Reuters to include in our report. All published data will be aggregated and summarized. Survey results will be fully anonymized for sharing and publication via one or more articles by project members. The results of our work will be broadly publicized by the PI, co-PI, and the entire team to its various communities of practice.

Reporting

The project will submit a Narrative Report written by the principal investigator with the support of the project team. The Narrative Report will review all project achievements, with expenditures included in the separate Financial Report. The structure of the Narrative Report will follow the reporting instructions from <http://msc.mellon.org/guidelines/reporting-instructions>. Since the project is just slightly over one year's duration, rather than producing an interim report at the end of project year 1 and a final report two months later, we propose to produce one complete report to summarize our research findings at the end of the project period: February 29, 2016, and one final Narrative and Financial Report summarizing the project per se, including a full accounting of the budget and expenses, in the 3-month period following the project's end date of May 31, 2016.

Appendices

Bibliography

- Achterkamp, Marjolein and Janita Vos 2007. Critically identifying stakeholders: Evaluating boundary critique as a vehicle for stakeholder identification. *Systems Research and Behavioral Science* 24(1): 3-14.
- Anderson, Kent 2013. 73 things publishers do (2013 edition). *Scholarly kitchen* ((Oct 22, 2013). <http://scholarlykitchen.sspnet.org/2013/10/22/updated-73-things-publishers-do-2013-edition/>.
- Aspesi, Claudio, Andrea Rosso, and Wielechowski 2012. Reed Elsevier: Transitioning to Open Access – Are the cost of savings sufficient to protect margins? Report. <http://www.richardpoynder.co.uk/OAcosts.pdf>.
- Babbie, Earl. *The Practice of Social Research*. 13th edition, Wadsworth, 2013;
- Báscones Dominguez, M. 2006. Economics of open access publishing. *Serials: The Journal for the Serials Community*, 19(1), 52-60.
- Bernard, Russell 2006. *Research methods in anthropology: Qualitative and quantitative approaches* (4th ed). Oxford: Alta Mira Press.
- Bijker, Wiebe, Thomas Hughes, and Trevor Pinch 1987. *The social construction of technological systems: New directions in the sociology and history of technology*. Cambridge: MIT Press.
- Björk, Bo-Christer and David Solomon 2014. Developing an Effective Market for Open Access Article Processing Charges. Report. http://www.wellcome.ac.uk/stellent/groups/corporatesite/@policy_communications/documents/web_document/wtp055910.pdf.
- Cook, Joel, Daniel Hulls, and David Jones 2011. Heading for the open road – costs & benefits of transitions in scholarly communications. A report for the JISC. <http://www.jisc.ac.uk/media/documents/publications/reports/2011/dynamicsoftransition.pdf>.
- Coulter, Gerry 2010. Launching and sustaining a scholarly journal on the internet: The International Journal of Baudrillard Studies. *Journal of Electronic Publishing* 13(1).
- Dallmeier-Tiessen, Suenje, et al. 2011. Highlights from the SOAP project survey. What Scientists Think about Open Access Publishing. arXiv:1101.5260.
- Davis, Philip M. and Cornell University Library. Task Force on Open Access Publishing 2004. Calculating the Cost per Article in the Current Subscription Model. (spreadsheet) <http://hdl.handle.net/1813/236>.
- 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.
- Edgar, Brian and John Willinsky 2010. A survey of scholarly journals using open journal systems. *Scholarly and research communication* 1(2).
- Finch Group, Accessibility, sustainability, excellence: how to expand access to research publications 2013. <http://www.researchinfonet.org/publish/finch/>.

Project Proposal for *Pay It Forward: Investigating a Sustainable Model of Open Access Article Processing Charges for Large North American Research Institutions*

Fisher, Julian 2008. Scholarly publishing re-invented: Real costs and real freedoms. *Journal of Electronic Publishing* 11(2).

Frass, Will, Jo Cross, and Victoria Gardner 2013. Open Access survey: Exploring the views of Taylor & Francis and Routledge authors. Report for Taylor & Francis Group and Routledge. <http://www.tandf.co.uk/journals/pdf/open-access-survey-march2013.pdf>.

Harley, Diane, Sophia Krzys Acord, Sarah Earl-Novell, Shannon Lawrence, and Judson King 2010. Assessing the future landscape of scholarly communication: An exploration of faculty values and needs in seven disciplines. Center for Studies in Higher Education, UC Berkeley. http://escholarship.org/uc/cshe_fsc.

Houghton, John, Bruce Rasmussen, Peter Sheehan, Charles Oppenheim, Anne Morris, Claire Creaser, Helen Greenwood, Mark Summers, and Adrian Gourlay 2009. Economic implications of scholarly publishing models: Exploring the costs and benefits. A report for the JISC. <http://www.jisc.ac.uk/media/documents/publications/rpconomicoapublishing.pdf>.

Mol, Annmarie 2002. *The body multiple: Ontology in medical practice*. Durham: Duke University Press.

Natriello, G and M. Rennick 2004. The middle way: The very low cost model for scholarly publishing. EdLab Research Publication. edlab.tc.columbia.edu/files/EdLab_Natriello4.pdf.

Odlyzko, Andrew 2013. Open Access, library and publisher competition, and the evolution of general commerce. Preprint. <http://www.dtc.umn.edu/~odlyzko/doc/libpubcomp.pdf>.

Oudshoorn, Nelly and Trevor Pinch 2003. *How users matter: The co-construction of users and technology*. Cambridge: MIT Press.

Pinfield, Stephen 2013. Medical research charities and open access. *Learned Publishing* 26(4): 285-302.

Research Information Network 2008. Activities, costs and funding flows in the scholarly communications system in the UK. Report.

Scholl, Hans 2004. Involving salient stakeholders: Beyond the technocratic view on change. *Action Research* 2(3): 277-304.

Shapiro, Lorna 2005. Establishing and publishing an online peer-reviewed journal: Action plan, resourcing, and costs. OJS project report. http://pkp.sfu.ca/files/OJS_Project_Report_Shapiro.pdf.

Sjostrom, Jonas and Goran Goldkuhl 2009. Socio-instrumental pragmatism in action. *Handbook of research on socio-technical design and social networking systems* (ed, Brian Whitworth and Aldo de Moor). Hershey, PA: IGI.

Smith, Scott M. Determining Sample Size: How to Ensure You Get the Correct Sample Size. Qualtrics E-Book, 2014. <http://success.qualtrics.com/rs/qualtrics/images/Determining-Sample-Size.pdf>

Solomon, David and Bo-Christer Björk 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.

Solomon, David and Bo-Christer Björk 2012. Publication fees in open access publishing: Sources of funding and factors influencing choice of journal. *Journal of the American Society for Information Science and Technology* 63(1): 98-107.

Project Proposal for *Pay It Forward: Investigating a Sustainable Model of Open Access Article Processing Charges for Large North American Research Institutions*

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

Suber, Peter 2012. *Open Access*. Cambridge: MIT Press.

Swan, Alma 2010. Modelling Scholarly Communication Options: Costs and benefits for universities. A report for the JISC. <http://eprints.ecs.soton.ac.uk/18584/>

Swan, Alma and John Houghton 2012. Going for Gold? The costs and benefits of Gold Open Access for UK research institutions: further economic modelling. Report to the UK Open Access Implementation Group.

http://repository.jisc.ac.uk/610/2/Modelling_Gold_Open_Access_for_institutions_-_final_draft3.pdf

Van Noorden, Richard 2013. Open access: The true cost of science publishing. *Nature* 495(7442): 426-429.

Walters, William H. 2007. Institutional journal costs in an open access environment. *Journal of the American Society for Information Science and Technology*, 58(1), 108-120.

Ware, Mark and Michael Mabe 2012. The stm report: An overview of scientific and scholarly publishing (3rd ed). Report. http://www.stm-assoc.org/2012_12_11_STM_Report_2012.pdf.

Waltham, Mary 2010. The future of scholarly publishing among social science and humanities associations. Report. <http://www.marywaltham.com/JSPfulltextarticleApril2010.pdf>.

Waltham, Mary; Carpenter, Todd, and Joseph, Heather 2004. A Survey of Business Trends at BioOne Publishing Partners and Its Implications for BioOne. http://www.marywaltham.com/Bio_one_article.pdf

Whitworth, Brian 2009. The social requirements of technical systems. *Handbook of research on socio-technical design and social networking systems* (ed, Brian Whitworth and Aldo de Moor). Hershey, PA: IGI.

Withey, Lynne, Steve Cohn, Ellen Faran, Michael Jensen, Garrett Kiely, Will Underwood, and Bruce Wilcox 2012. Sustaining scholarly publishing: New business models for university presses. Report. <http://www.aaupnet.org/images/stories/documents/aaupbusinessmodels2011.pdf>

University of California Regents, 2014

- Richard C. Blum, Chairman of Blum Capital Partners, L.P.; Co-Chairman of Newbridge Capital, LLC.
- William De La Peña, M.D., Ophthalmologist and medical director of the De La Pena Eye Clinic.
- Russell Gould, Partner, California Strategies LLC.
- Eddie Island, retired attorney and executive.
- George Kieffer, Partner and member of the Executive Committee of the national law firm of Manatt, Phelps & Phillips, LLP.
- Sherry L. Lansing, Founder of the Sherry Lansing Foundation and former Chair and CEO of Paramount Pictures' Motion Picture Group.
- Monica Lozano, Chief Executive Officer and Chair of the Board of Impremedia, LLC.
- Hadi Makarechian, Chairman of Makar Properties Board of Directors.
- Norman J. Pattiz, Founder and former Chairman of the Board of Westwood One and CEO of Courtside Entertainment Group.
- Bonnie Reiss, Global Director of the Schwarzenegger Institute for State and Global Policy at the University of Southern California.
- Fred Ruiz, Co-founder and Chairman Emeritus of Ruiz Foods.
- Sadia Saifuddin, Student at University of California, Berkeley.
- Richard Sherman, Chief Executive Officer at the Geffen Company.
- Bruce D. Varner, Partner in the law firm Varner & Brandt.
- Paul Wachter, President and CEO, Main Street Advisors.
- Charlene Zettel, Chief Executive Officer, Donate Life California