

Discussion of a Bottom Up Approach to Examining Economic Impacts of Open Access

Don King
(donalddwking@gmail.com)

This discussion follows on John Houghton's excellent preliminary paper on "Economic impacts of open access: Some preliminary thoughts on scoping a proposal". He presents a publishing business model (from a forthcoming report by he and Colin Steele) that distinguishes four hybrids to the current subscription model and two widely recognized open access models: (1) archives and institutional repositories and (2) author/institution pays. Without quite emphasizing it, the paper puts the impacts into a systems framework with a focus on a top down perspective. At the meeting I suggested a complementary bottom up approach which is the basis of this discussion.

Both systems approaches would rely on the overall goals of the journal communication system currently in place. Examples of goals might be:

- Improve worldwide research, health, education, etc.
- Maximize economic and social returns to public investment in R&D
- Ensure that all potential users have unhindered access to needed information in their required language, format, and other attributes

Comparison of open access models with the current system should be made in light of achieving these goals. John Houghton and Peter Sheehan described the second possible goal above and presented a simulation for assessing the goal. Further discussion at the meeting began to outline possible "goals".

The Houghton paper suggests some possible benefits of open access which lead to such "goals". For example, (1) improved speed of access leads to speeding up research and discovery process, etc. and (2) improved access leads to less duplicate research, etc. To illustrate a bottom up approach, consider the example above. I believe that it is necessary to measure current speed of access and its effect on speed of research and discovery process as a first step and then carefully examine how open access can lead to even faster research and discovery.

Speed of access could mean reducing the duration of time between achieving research results and documentation and when it is made available. In science, Carol Tenopir and I and our colleagues have observed surveys of scientists in which article reading is sometimes said to speed research (although we did not attempt to determine by how much). Furthermore, we have observed the age of articles read for research (and other) purposes. Some of these articles are read immediately upon receipt from publishers. From these two results one can determine how many of the recently published articles read result in speeding research. Some of these new articles might have speeded research more had open access provided earlier access. The issue

then is to describe how earlier access is provided through open access, observe whether earlier access occurs, and compare open access results with the current results.

Speed of access and other such benefits are dependent on the entire journal communication system of functions, participants who perform the functions, resources used by the participants and others who influence the system such as funders. A comparison of the current system with open access (by either model) needs to be made for each system function, what participants are necessary to perform the functions, and what resources are required.

Below are a few top-of-the-head examples of how one might characterize these aspects of the journal communication system and how open access would compare with the current system. The examples are not necessarily the most important ones, but serve to illustrate the kind of initial and comprehensive thinking that might be done. It seems to me that a first step in any comprehensive study is to think through these details starting with common communication functions such as funding research; performing the research; documenting the research results; reviewing/refereeing the documented results; editing the information contents; translating information content; synthesizing the information for search and retrieval; communicating the information contents to readers or to intermediaries who can store/archive and provide access to the information at a time when needed; searching, locating, and obtaining the information content; and user screening/selection, reading and the actual use of the information (for many purposes).

Once the basic communication functions are identified and described, current participants who perform the functions can be described and potential open access participants determined. Some functions and participants may be common to the current and open access systems. For example, in the US the National Science Foundation specifies a matrix of classes of R&D funders and performers in name and expenditures. These funder and performer participants are not likely to change due to introduction of open access. On the other hand, some functions such as searching, locating, and obtaining information content could change appreciably with open access for some participant users. For example, we have observed in the past that most small, high tech firms in the U.S. do not have adequate library capabilities (due in part to lack of library economies of scale). For this reason, many of these firms rely on academic or public libraries to gain access to the journal literature at cost to them in speed of access defined by the duration of time between when information is needed and when it can be accessed. Reliance on academic and public libraries also has a significant affect on scientists' time and effort. These consequences of the current system can be measured. That is, we have evidence of the extent of use of public and academic libraries by staff of small businesses and how much it costs them in their time to use a public or academic library. Presumably, open access would provide nearly all needed information to these firms at comparatively increased speed of access and reduced effort. The same concept would apply to researchers in developing countries.

In addition to participants, it is useful to examine resources required by current participants and compare the amount and cost of these resources to those required for open access. By resources, I mean people, equipment, systems, facilities, supplies (e.g. paper), etc. Changes in peoples' time is crucial because reduction in time can be reallocated to better use and scientists' time dominates the cost of many functions. In the example above where open access serves small

businesses, the time required by researchers to visit academic and public libraries can be reallocated to their research. Another example involves smaller academic and public libraries that depend heavily on interlibrary borrowing (or document delivery) and larger libraries that do most of the interlibrary lending. These activities currently require a substantial amount of library staff time, as well as, space, systems, and equipment. Presumably open access can reduce or perhaps eliminate the need for this library activity. Our cost studies and others as well provide evidence of the resources expended in interlibrary lending and borrowing and, when extrapolated to all lending and borrowing, provide measures to compare with open access since these library activities would essentially be eliminated. We also have some evidence of the cost to obtain information through interlibrary borrowing compared with direct access. Furthermore, these activities affect access time in the second sense above.

There are other aspects to comparing open access with current journal communication. One aspect has to do with the monetary cost of participant resources and comparison of current and open access costs. A case in point is the cost of library print and electronic journal collections. Most emphasis of library cost has dwelled on the price paid for collections in subscription price and license fees. Yet the cost of negotiating, acquiring, processing, using and maintaining the collections can be appreciable. For example, at the University of Pittsburgh the overall cost of resources used to perform the functions above are: collection purchase price (51% of total cost); staff (44%); space (2%); binding (1%); photocopying (0.2%); staff workstations (0.4%); user workstations (0.1%); and systems (0.4%). This pattern may not be typical because the Pittsburgh library was in transition at the time of our cost study (2003) with heavy duplication of the two collections. However, the results do indicate that open access might well affect the cost of some library resources beyond the collection price paid.

Similarly, publishing costs should be re-examined under current and proposed open access models, both in print and electronic format because most personal subscriptions and some library collections continue to be read in print. Some have reported that article processing costs less with open access than with current practice. However, it is difficult to visualize how open access by author/institution pays might appreciably affect manuscript receipt processing, initial disposition decision-making, identifying reviewers or referees, review processing, subject editing, special graphic and other preparation, formatting, copy editing, processing author approval, indexing, coding, redaction, or preparation of master images. Some activities might be dropped or done differently, but that could be the case for either current or author/institution pays model. Others have quoted relatively low author/institution pay fees (e.g., PLoS) as a basis for suggesting that open access publishing costs less. However, the lower article processing costs of this model could just as easily be adopted for the current model. That is, there is no apparent reason why the PLoS journal articles could not be traditionally published at a lower price (reflecting their low cost) than most current publications. If open access leads to new ways of distribution or of providing access, then examining these comparative costs makes sense. However, the alternative processes need to be carefully defined and described.

A companion aspect to participant cost is the flow or transfer of funds among participants. Some attention has been paid to this aspect by speculating about how open access might affect funds to and from libraries (and their funders) and publishers, but little thought given to the effect of open access on the entire system of participants. Again, I would start with a detailed picture of

the flow of funds among the current journal communication system participants starting with where support comes to R&D funders and other sources of research funding of the researchers. A few other examples of types of flow or transfer of funds might be:

- From R&D funders (i.e., government, industry, foundations, universities, etc.) to research performers (i.e., government, industry, universities, small high tech firms, etc.)
- From R&D performers to scientists (in salaries)
- From scientist to government (in taxes)
- From scientists to publishers (to pay as individuals for personal subscriptions, author fees, reprints)*
- From university sources of funds (i.e., government, student tuition, endowments, non-research grants, etc.) to the universities
- From university and other organization funders to their operations (i.e., libraries, administration, departments, etc.)
- From operations to non-library journal-related facilities, systems and equipment, scientists as authors, other related staff, etc.*
- From libraries to publishers (to pay for subscriptions or licenses)*
- From libraries to journal-related intermediaries and vendors*
- From universities and other organizations to publishers (to pay for some personal subscriptions, author fees, preprints, etc.)*
- From intermediaries and vendors to publishers
- From intermediaries and vendors to staff (in salaries) and other journal-related resources*
- From intermediary and vendor staff to government (in taxes)*
- From publishers to staff outside editors and referees (sometimes) and other resources*
- From publisher staff, editors and referees to government (in taxes)*
- From publishers to lenders*
- From publishers to investors*
- From publisher profit to government (in taxes)*
- From investors to government (in taxes)*
- From government to R&D and other funding

Some of this transfer of funds may be trivial and some important transfers may not have been considered above. The transfers with asterisks indicate those that might be affected by open access.

Analysis of such flow of funds could be like an input-output model Analysis. I suspect that such an analysis might reveal some unexpected results. Sometimes detail is useful. For example, universities as R&D performers rely on many funders in the US such as federal government (58% of funds), universities (20%), other non-profit (8%), industry (7%), and other (7%). Yet, the amount received from funders might not reflect the number or proportion of articles yielded from the funds. As a case in point, we recently queried science journal authors at the University of Pittsburgh as to where they received funding for research leading to articles. The principal sources were from the University itself (35% of articles), government (33%), industry (25%) and with foundations following (7%). The point is that this analysis should be generalized so that one can assess the effect of the funders' paying author fees; recognizing that the amount of payment could well detract from research being funded just as the time spent writing articles (about 80 to 100 hours on average) detracts from time that could have been spent on research. I am not against either, but feel that we must thoroughly understand what is really happening. [In fact, I have often written about the public good aspects of government sponsored technical documents and the value of facilitating their maximum use].

One often mentioned lament is that universities pay for research leading to journal manuscripts (information) that are given to publishers who in turn sell the articles (information) back to their libraries. Of course publishers argue that they provide two services: (1) value-added processing to information content and (2) distribution or access to those who need the information. A complete analysis of the amount and flow of funds would show where the research funding (leading to articles) comes from, where all publishing funding/revenue comes from (including subscriptions, advertising, association subsidation, etc.), and where library funding comes from (within universities and within other types of libraries). Subscription revenue comes not only from university libraries but also from public and special libraries. Personal subscriptions are sometimes paid by scientists themselves, by grants, through society memberships, and by non-university parent organizations. The point is that a better characterization of this flow of funds and how open access would affect this flow is an essential component of the economic impacts of open access.

Another important component of the flow of funds deals with journal financing, whether profit or non-profit. There are several aspects of financing journals. First there is a substantial cost to starting a new journal (i.e., \$100,000 is a commonly quoted amount). Therefore, there must be a source of funds to start a journal and to keep it going (often at a loss) until it can sustain itself and provide a return on this initial investment. This "unprofitable" period of time can be anywhere from 5 to 10 years. Nature is said to have taken 40 years. Of course, SPARC and PLoS have received funds for starting new journals from foundations. Societies often support the start of new journals, but less often in recent years. Publishers with large portfolios sometimes have "profitable" journals subsidize new journals and some popular, but "unprofitable", ones. Commercial publishers rely on stockholder investment to start new journals. Here again an accurate current portrayal is useful to provide a comparison to open access.

There are other requirements for investment as well. For example, some costs (and payment) are incurred before revenue is received. Therefore, some investment or borrowed funds are necessary to cover these costs which tend to be labor-intensive and must be paid in a timely

manner. Another reason for investment or borrowed funds is to support capital investment and publishing R&D. Larger publishers can afford new equipment, facilities, etc., and R&D expenditures. Examination of flow of funds should determine how much R&D funding flows from government (e.g., NSF) or foundations to societies or other non-profit publishers. At one time the amount from NSF was substantial in the U.S.

Commercial publishers have been criticized that they receive investment funds from investors who in turn receive a portion of profit as a return on their investment. However, from a flow of funds perspective, one can argue that stockholder investment funds do not detract from funds that could be applied to science research or communication. If a new journal requires a \$100,000 investment, it is in some ways better to have the funds come from non-science sources of investment than from societies, government, or foundations that could invest the \$100,000 in science research. Excessive profit, of course, appears to be unseemly, but gross profit is applied to taxes (about 50% in the U.S.), paying back loans, R&D and capital requirements, as well as the remaining distributed to stockholders who also pay taxes on capital gains. There are arguments against commercial publishing, but the matter is not as straight forward as might be thought.

Summary

I am not suggesting that a bottom up approach is better than a top down one. I believe that both approaches contribute substantially to an understanding of the journal communication system and the potential economic impacts of open access (both positive as well as negative).

As I mentioned at the meeting, it seems wise for someone (not me due to excessive commitments) to think through a bottom up approach to compare open access models with the current journal communication system of functions, participants, resources, costs and flow of funds. It is tedious work, but should highlight where economic impacts of open access can make the most inroads. John Houghton and others have considered components of the system, but more work needs to be done. Perhaps the Finnish/EU project serves as an initial approach. I've tried to give a few simple examples from research my colleagues and I have conducted, but these examples merely scratch the surface. There are many other examples to be drawn from and much research needs to be new.

As a statistician who has conducted many surveys for government and others, it has been my role to observe and report. I've tried to avoid adversarial positions such as between libraries and publishers. Certainly, this was essential in the government study I did in the 1970s to determine to consequences of the 1976 Copyright revision. It was during this study that I realized the rift between libraries and publishers was based largely on faulty assumptions and preconceived notions. Publishers, in particular, were convinced that libraries were "ripping them off" through Interlibrary Lending and that their revenue was being seriously compromised. In negotiations the publishers permitted a "fair use" provision that borrowing libraries could borrow five or fewer articles from a journal without royalty payment. This provision essentially finessed the royalty payment issue because, at the time, five uses of a journal was typically a breakeven point below which it costs less to borrow and above which it costs less to subscribe. Several library cost studies had concluded this phenomenon prior to Copyright revision. Our subsequent

survey clearly indicated that this was the case and history (and a study I did later for CCC) showed that publisher revenue was not significantly affected by ILL.

I continue to believe that the future of the journal communication system depends on cooperation among all system participants. One motivation of the book authored by Carol Tenopir and me was to provide evidence of the contribution made by all participants, to describe how they operate and why they make decisions they do, and to emphasize their mutual interdependence. My one advocacy position is the furtherance of science and I continue to believe all participants have this goal in mind, at least secondarily so because their survival depends on it.