

Scientific journals, overlays and repositories: a case study of costs and sustainability issues

Panayiota Polydoratou and Martin Moyle

University College London,
Library Services,
DMS Watson Building, Malet Place, WC1E 6BT
lib-rioja@ucl.ac.uk

Abstract. Publishing scientific research is an area of study that attracts interest from various stakeholders such as publishers, academic and research staff, libraries and funders. In the past decade increased journal subscription fees prompted calls for cheaper and more efficient means of accessing the scientific literature. Factors such as the expansion of digital repositories, the introduction of open source journal management software, an increasing awareness within the scholarly community at large of the issues around open access, and an increasing readiness within the publishing community to experiment with new models, suggest that the circumstances may now be right for new models of scientific publishing to be explored, as well as potential business models and sustainable solutions around them. This paper explores some of the issues around the costs and sustainability of a prospective journal model known as the overlay journal. We present estimates of initial start up costs for such a model, discuss the factors that would influence scientists in deciding whether to publish in a journal overlaid onto a public repository; and report their views on the relative importance of different features and functions of a journal in terms of funding priorities.

Introduction

This paper presents findings from the JISC (Joint Information Systems Committee, UK) funded RIOJA (Repository Interface for Overlaid Journal Archives) RIOJA (Repository Interface for Overlaid Journal Archives) project (<http://www.ucl.ac.uk/ls/rioja>) which aimed to address the issues around the development and implementation of a new publishing model, the overlay journal. For the purposes of this paper, an overlay journal is defined as an open access journal whose content is deposited to and resides in one or more open access repositories. It utilizes quality certification, is sustainable and adheres to preservation standards.

The impetus for the RIOJA project came directly from academic users of the arXiv (<http://arxiv.org>) subject repository. For this reason, arXiv and its community is the

testbed for RIOJA. arXiv was founded in 1991 to facilitate the exchange of pre-prints between physicists. It now holds over 495,000 scientific papers, and in recent years its coverage has extended to mathematics, nonlinear sciences, quantitative biology and computer science in addition to physics. arXiv is firmly embedded in the research workflows of these communities.

Background and statement of the problem

The overlay concept, and the term "overlay journal" itself, appear to be attributed to Ginsparg (1996). Significant contribution to the concept of overlay journals has been conducted by J W T Smith (1999) who discussed and compared functions of the existing publishing model what he referred to as the "deconstructed journal".

Halliday and Oppenheim (1999), in a report regarding the economics of Digital Libraries, recommended further research, in the field of electronic publishing in particular. Specifically, they suggested that the costs of electronic journal services should be further investigated, and commented that the degree of functionality that users require from electronic journals may have an impact on their costs. In a JISC funded report, consultants from Rightscom Ltd (2005) suggested that commercial arrangements for the provision of access to the published literature are made based on the nature of the resource and the anticipated usage of the resource. Cockerill (2006) indicated that what is regarded as a sustainable publishing model in the traditional sense (pay for access) is actually supported by the willingness of libraries to pay [... "even reluctantly", p.94] large amounts of money to ensure access to the published literature. He suggested that as open access does not introduce any new costs there should not be any difficulty, in theory, in sustaining open access to the literature. Waltham (2005) raised further questions about the role of learned societies as publishers as well as the overall acceptance of the 'author pays' model by the scientific community.

Self-archiving and open access journals have been recommended by the Budapest Open Access Initiative (<http://www.soros.org/openaccess/read.shtml>) as the means to achieve access to publicly-funded research. The overlay model has the potential to combine both these "Green" (self-archiving) and "Gold" (open access journal) roads to open access. Hagemmann (2006) noted that "...overlay journals complement the original BOAI dual strategy for achieving Open Access..." and suggested that the overlay model could be the next step to open access. In support of open access to information the BOAI published guides and handbooks on best practice to launching a new open access journal, converting an existing journal to open access, and business models to take into consideration [Crow & Goldstein, 2003a-c).

Factors such as the expansion of digital repositories, the introduction of open source journal management software, an increasing awareness within the scholarly community at large of the issues around open access, and an increasing readiness within the publishing community to experiment with new models, suggest that the

circumstances may now be right for an overlay model to succeed. Part of the RIOJA project was to test the reaction of one research community, selected for its close integration with a central subject repository, to this prospective new model.

arXiv and the publishing process

Despite the everyday importance of arXiv to researchers, depositing papers to the repository remains a supplement to the traditional publishing process, rather than a replacement for it. Peer review is as important to arXiv-depositing researchers as to those in other scientific disciplines, and, to achieve peer acceptance, papers continue to be submitted for publication in the traditional way. Once a paper is accepted for publication, an author will typically update the corresponding arXiv version to denote the publishing journal title and the date of acceptance. These annotations, indicating acceptance for publication, serve as badges of quality for arXiv deposits.

Prosser (2005), quoting GermanMan (199x) who must get credit, notes that journals are traditionally held to perform four "first order" functions:

- Registration: an author wishes to be acknowledged as the person who carried out a specific piece of research and made a specific discovery
- Certification: the author's claims are tested through independent peer review, and it is determined that they are reasonable
- Awareness: the research is communicated to the author's peer group
- Archiving: the research is retained for posterity

It is clear that arXiv already provides three of these functions:

Registration occurs when a research paper is received by arXiv, at which point it is assigned a unique identifier and date stamp. It is commonplace for papers to be cited thereafter by arXiv reference number, illustrating the acceptance of the arXiv registration process.

Once registered, a paper can appear in the public domain on the same day. It is openly and freely available, without barriers to access. arXiv also offers email alerting to new papers and is compliant with OAI-PMH. It fulfils the *Awareness* function: many researchers clearly consult the repository in preference to traditional journals.

arXiv also satisfies the *Archiving* function, with an emphasis on stable and portable formats at ingest, and the retention for public scrutiny of version-controlled superseded papers alongside the most recent update.

arXiv, therefore, provides three of the four "first order" functions of the traditional journal. It does not yet provide *Certification*. To achieve a quality stamp, researchers from arXiv's subject communities and their institutions must engage with the full, protracted and costly machinery of formal publication. This can involve delays, page charges, author/funder charges, restrictive copyright transfer agreements, version control issues between the arXiv holdings of a paper and its published counterpart, and post-publication barriers to access because of subscription and licensing arrangements; and yet the content of the resulting journal productions has often

already been disseminated via arXiv and consumed by researchers. During the development and implementation of the RIOJA tool (see below) we were able to estimate some initial start up costs which alongside the surveys' findings allowed to draw some cost projections for the overlay journal model.

Methodology

This paper builds on the results from two community surveys which were undertaken to explore the views of scientists in the fields of astrophysics and cosmology concerning the feasibility of an overlay journal model. The community surveys comprised of:

- An online questionnaire survey targeting more than 4000 scientists from the top 100 universities and 15 non academic institutions in science (yielded response by 683 scientists, 17% response rate),
- Interviews with publishers and members of editorial boards of peer-reviewed journals. These complementary studies were intended to enable a more rounded understanding of the publishing process, and to help the project to explore whether an overlay journal model in astrophysics and cosmology could be viable in the long term.¹

In addition, the authors undertook desktop research to identify studies on the costs of publishing scientific journals and to compare, where possible, their findings against the development and implementation of the RIOJA toolkit described below.

The RIOJA toolkit

The technical part of the project dealt with the development of XML-based APIs for the exchange of data between digital repositories and journals to facilitate the overlay of academic journals onto separate digital repositories. It was assumed that: a) the repository provides the registration, awareness and archiving functions of a journal and b) the journal provides only the certification (peer review) and additional awareness functions. All versions of a paper are stored in the repository, from the original submission to the published version and beyond. The repository can tag papers with their status, so end users can, if desired, filter papers to see only submitted, accepted or published papers as they prefer. The journal tracks different versions of the repository paper, and applies its final "published" quality stamp to one particular "final" paper version. The repository may, however, allow updates to a paper after publication, allowing easy access to a corrected version as well as the "published" version. The APIs are implemented in the RIOJA project's test bed, and (partially) in the arXiv subject repository (Lewis, 2007).

¹ Published results from the RIOJA project community surveys can be found at <http://www.ucl.ac.uk/ls/rioja/dissemin/>

Initial start up costs

The RIOJA toolkit saw the development of a module specification to support automated interactions with repositories. In full, the technical work comprised:

- Development of open API for communication between repositories and journals
- Development of software for hosting overlaid journals using the API
- Demonstration journal software, using the RIOJA API implemented on arXiv.org repository
- Version of ePrints repository software to incorporate RIOJA APIs for application in any subject area (N.B. still in progress)

Start-up costs included the fee to the company to which the development was outsourced. Overall, initial development and implementation costs, excluding person power, did not exceed £5000 (\$7500).² Indicative amounts are listed below:

- Software development costs ~ 4000 (\$6500)
- OJS developed LaTeX plugin ~200 (\$400)
- Web hosting ~200/per year (\$400)

Fixed and variable costs

The term fixed costs is used in the literature to identify those costs associated with the publishing process that remain the same regardless of circulation (King, 2007; SQW Limited 2004). By contrast, variable costs refer to those that change with the number of subscriptions (e.g. cost of reproduction, subscription maintenance, etc.). Some of these costs are associated with particular business models (e.g. subscription based model) and publishing media (e.g. print versus electronic) which raises the question as to whether a cost recovery model such as the “author pays” could be cheaper to sustain in an electronic environment and using the overlay journal as a model.

Some of the costs referring to the registration stage concern submission. Costs at submission level include both rejected and accepted papers, are in general fixed costs and include what is usually addressed as first copy costs. Those include costs linked to article processing such as the work of the editor and editorial board, system support (administrative and managerial aspects, the organisation of the peer review process, staff involved in the system, etc). In those costs should be included those that refer to non-article processing. The average cost of first copy production varies widely in different sciences. King (2007) presents findings from previously reported first copy costs, ranging from \$450 to \$2500 for article processing and reaching to \$10000 in some disciplines. Consultants in SQW Limited (2004) reported that first copy costs for a good to high quality journal are estimated at around \$1500 (\$1650 including first copy and fixed costs). However, distribution costs do not vary with the number of subscriptions and are in the majority fixed rather than variable. Furthermore, it is

² Exchange rate of 1 GBP = 1.98 USD (13/08/2008)

even easier to separate and control submission costs if a submission fee and a publication fee is set separately.

Community uptake

The community surveys received responses from 683 scientists (17% of 4012 contacted), and representatives from publishing houses and members of editorial boards from peer-reviewed journals in astrophysics and cosmology. Results indicated that more than half of the respondents (53%) were favourably disposed to the idea of overlay journal as a potential future model for scientific publishing. Over three quarters (80%) of the respondents were, in principle, willing to act as referees in an arXiv-overlay journal.

The most important factors which would encourage publication in a repository-overlaid journal were the quality of other submitted papers (526 responses), the transparency of the peer review process (410) and the reputation of the editorial board (386). Respondents also provided a range of other factors that they considered important, among them the reputation of the journal; its competitiveness measured against other journals under the RAE (the UK's Research Assessment Exercise); the quality both of the journal's referees and of its accepted papers; a commitment to using free software; a commitment to the long-term archiving and preservation of published papers; relevant readership; and its impact factor, (which, it was noted, should only take into account citations to papers after final acceptance and not while residing on arXiv prior to "publication").

The interviews with publishers and editors did not reveal any substantial information about costings that have not already been reported in the literature (King, 2007, SQW Limited, 2004; Waltham, 2004) or are available on some publishers' websites, e.g. PhysMath Central (<http://www.biomedcentral.com/info/about/apcfaq>). Interviewees suggested that the processing price per article varies by journal, discipline and usage. However, it was noted that community uptake and in particular the interest of academic and research staff in new publishing models is the prime driver for their adapting to technology challenges. For example, one of the publishers interviewed stated that one of their most successful journals, both in terms of revenue to the publisher and in terms of perceived quality and acceptance by the scientific community, was converted to open access (the 'author pays' model) purely because of community demand.

Journal functions

Meanwhile, a question included in the questionnaire survey concerning how expenditure should be apportioned towards particular functions of a journal was subject to criticism: respondents queried whether a scientist has adequate knowledge of the publishing process and its associated costs to make any useful observations. It was also observed that the publishing process entails more than the distribution phase, which some respondents felt that the survey, and by implication the overlay model, appeared only to address. However, the costs associated with the work of scientific

editors, with the integrity and long-term archiving of journal content, and with the transparency of peer review were highlighted as worthwhile (Table 1, scale 1 (little) – 5 (most of the amount)). An indicative comment is reproduced below:

“... Very-little of a high-cost journal may be more than a considerable amou[n]t of a low-cost one. Perhaps it would be better posed in terms of one's priorities in paying for the journal. I think that in this day paying those such as the editors and referees, and ensuring the integrity of the archive, ought to be a higher priority than producing a paper version of the journal. Especially for an overlay journal such as you propose”.

Suggested expenditure/priority	None	1	2	3	4	5	Not sure
Paying scientific editors	23	23	60	246	141	15	21
Paying copy editors	8	28	73	256	134	6	15
Maintenance of journal software	4	20	73	238	147	9	30
Journal website	5	28	79	225	149	20	15
Online archive of journal's own back issues	9	27	52	202	189	18	19
Production of paper version	138	101	125	107	29	4	14
Extra features such as storage of associated data	30	63	105	182	100	6	26
Publisher profits	142	122	138	91	9	0	19
Paying referees	249	70	70	85	22	8	18
Other	3	1	1	1	3	2	3

Table 1. Suggested expenditure/priorities

Copy editing

Copy editing, the level of author involvement in it, and who should be responsible for any costs associated with it, were also issues that were commented upon. Some respondents favoured the idea of charging extra for papers that require extensive copy editing. Almost half of the respondents favoured the suggestion that the cost of copy editing should be borne by the author, and that it should also be variable based on the amount of copy editing required. Furthermore, almost half of the respondents (47%) appear to be in agreement that those changes should be carried out by the author (Table 2). The appearance and layout of the published papers were considered important.

“The idea of charging authors for papers that require excessive copyediting is a great one!”

“Copy editing is a difficult issue: it should be the [responsibility] of authors to improve their writing, on the other hand the journal should take [responsibility] for what it published. Perhaps an author could have say three chances and after that should pay for copy editing?”

“...my position is that a basic copy editing should be provided by the journal, but that extremely messy papers should be penalized, perhaps by introducing extra costs”

“I do believe money [is] being wasted on the copy-editing of already copy-edited articles, on paper copies of journals, on library subscriptions, etc. The publications process needs to be streamlined and a new type of open-access peer-reviewed journal might just be the right thing”.

Statement	Rating	% agree	95% confidence limit
The cost of copy editing should be borne by the author and vary from paper to paper, depending on the amount of copy editing required		48.2	± 3.8
Copy editing should be carried out by the author		47.3	± 3.8
A referee should be prepared to assess whether or not copy editing is required		18.1	± 2.9
The cost of copy editing should be borne by the journal		11.1	± 2.4
When a journal makes copy edits, the corrected LaTeX should be returned to the author (after his/her approval)		4.7	± 1.6

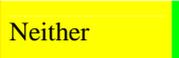
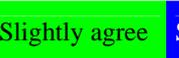
Key:  Strongly disagree  Slightly disagree  Neither  Slightly agree  Strongly agree

Table 2. Copy editing

Funding

When asked where the funding to meet those costs should come from, the respondents preferred to select research funders (485 people, 71% of base=683), library subscriptions (432 people, 63%) and sponsorship, for example by a Learned Society (350 people, 51%). Models requiring an author to pay from research funds either on acceptance (218 people) or on submission (47 people) of a paper were not endorsed. Other possible funding sources mentioned in comments included: personal donations, professional association contributions, commercial and/or not-for-profit organisations, advertisements, subscriptions and even models in which authors pay partially on submission and partially on acceptance.

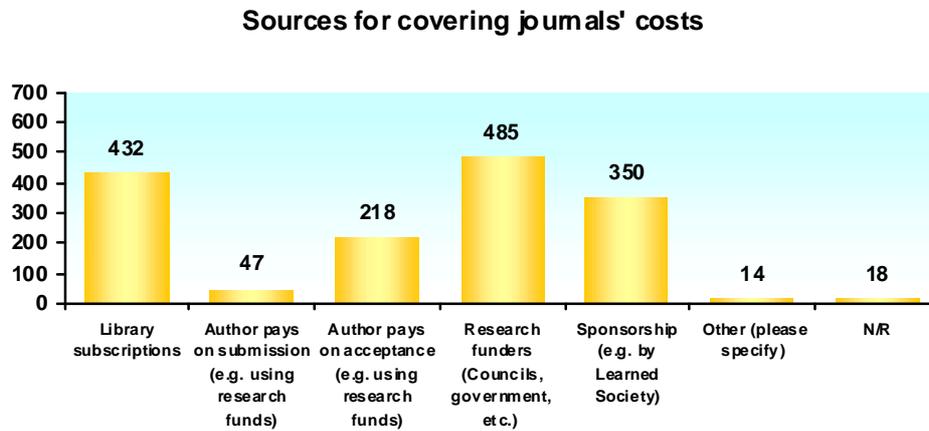


Fig. 1. Sources for covering journals' costs

Summary and conclusions

The RIOJA community surveys found some encouragement for the overlay journal model in the fields of Astrophysics and Cosmology. However, they raised several implementation issues that they would consider important, primarily relating to the quality of the editorial board and of the published papers, and to the long-term archiving of the accepted research material. The traditional copy-editing function remains important to researchers in these disciplines, as is visibility in indexing services. The traditional printed volume is of little interest. These are generic concerns, for which repository overlay is not necessarily the complete answer.

Although the interviews with publishers and editors did not reveal any substantial information about costings that have not already been reported in the literature (King, 2007, SQW Limited, 2004; Waltham, 2004) or are available on some publishers' websites, some of the cost projections and business analysis for the development and maintenance of a journal founded on overlay certification in this field could help to inform future undertakings of this nature in different disciplines and with different repositories. Case studies or scenarios which involve setting up a new journal or converting an existing one to an overlay model will allow more precise definition of costing projections..

References

- Cockerill, M. Business models in open access publishing in, *Jacobs, Neil (ed.) Open Access: Key Strategic, Technical and Economic Aspect*, Oxford: Chandos Publishing, pp. 89-95, 2006. Available at: <http://demo.openrepository.com/demo/handle/2384/2367> (Last accessed 13/08/2008)
- Crow, R. & Goldstein, H. 2003a, Model Business Plan: A Supplemental Guide for Open Access Journal Developers & Publishers, Open Society Initiative. Available at: http://www.soros.org/openaccess/oajguides/oaj_supplement_0703.pdf (Last accessed 13/08/2008)
- Crow, R. & Goldstein, H. 2003b. Guide to Business Planning for Launching a New Open Access Journal, Open Society Institute. 2nd edition. Available at: http://www.soros.org/openaccess/oajguides/business_planning.pdf (Last accessed 13/08/2008)
- Crow, R. & Goldstein, H. 2003c, Guide to Business Planning for Converting a Subscription-based Journal to Open Access, Open Society Institute. Available at: http://www.soros.org/openaccess/oajguides/business_converting.pdf (Last accessed 13/08/2008)
- Ginsparg, P. (1996). Winners and Losers in the Global Research Village. Invited contribution, UNESCO Conference HQ, Paris, 19-23 Feb 1996. Available at: <http://xxx.lanl.gov/blurb/pg96unesco.html> (Last accessed 13/08/2008)
- Haggemann, M. (2006). SPARC Innovator: December 2006. Available at: <http://www.arl.org/sparc/innovator/hagemann.html> (Last accessed 13/08/2008)
- Halliday, L and C Oppenheim. (1999). Economic models of the Digital Library. Available at: <http://www.ukoln.ac.uk/services/elib/papers/ukoln/emod-diglib/final-report.pdf> (Last accessed 13/08/2008)
- King, D. (2007). The cost of journal publishing: a literature review and commentary. *Learned Publishing*, Vol. 20 (2), pp. 85-106
- Lewis, A. 2007. RIOJA Journal-Repository APIs. Available at: <http://cosmologist.info/xml/APIs.html> (Last accessed 13/08/2008)
- Prosser, David C. (2005) *Fulfilling the promise of scholarly communication - a comparison between old and new access models*, in Nielsen, Erland Kolding and Saur, Klaus G. and Ceynowa, Klaus, Eds. *Die innovative Bibliothek : Elmar Mittler zum 65. Geburtstag*, pp. 95-106. K G Saur. (Also available at <http://eprints.rclis.org/archive/00003918>) (Last accessed 13/08/2008)
- Rightcom Ltd. Business model for journal content: final report, JISC. 2005. Available from : http://www.nesli2.ac.uk/JBM_o_20050401Final_report_redacted_for_publication.pdf (Last accessed 13/08/2008)
- Smith, J W T. (1999). The deconstructed journal: a new model for academic publishing. *Learned Publishing*, Vol. 12 (2), pp. 79-91
- SQW Limited 2004, Costs and business models in scientific research publishing: a report commissioned by the Wellcome Trust, Wellcome Trust
- Waltham, M. *Learned Society Open Access Business Models*, JISC. 2005 Available at: http://www.jisc.ac.uk/uploaded_documents/Learned%20Society%20Open%20Access%20Business%20Models.doc (Last accessed 13/08/2008)