RIOJA

**Repository Interface for Overlaid Journal Archives:
costs estimates and sustainability issues**

Panayiota Polydoratou

July 2008

# Contents

## *Introduction*

The RIOJA project ([http://www.ucl.ac.uk/ls/rioja](http://www.ucl.ac.uk/ls/rioja)) investigated the feasibility of an overlay journal model in collaboration with the arXiv and in the scientific domain of astrophysics and cosmology. Scientists in this community are active users of e-prints repositories such as the arXiv. Furthermore, they have the support of Professional Associations and Learned Societies that have been pioneers in adapting to new publishing models and in particular electronic journals.

Long term access to information as well as maintaining provision to sustainable systems/services is important to various parties in the scholarly communication system: the creators of information, developers and managers of services, libraries, publishers, funders and also users. Although scientific journals have been in existence since the 18th century (Lawal, 2001), factors such as increased journal subscription prices in the last decades and the emergence of new technologies have triggered discussions on the potential of new business models for publishing research. Furthermore, the advent of the open access movement also contributed to exploration of the issues around free access to information and provision of sustainable services. Exploring aspects of sustainability is something that should be seen over a period of time and whether launching, converting or simply maintaining a new or existing system/service the needs of the community it serves should be taken into account.

Scientific journal publishing is a complex process. Besides disseminating scientific knowledge, registration of a claim for new discovery and a quality "stamp" it also facilitates social factors. Besides making research findings available and contributing to the advancement of knowledge, publishing is also a means for measuring quality of the work of scientists, allocating funding, and acknowledging contributions to knowledge.

In this report, we will try to provide an overview of a new publishing model, that of the overlay journal. We will discuss the use of the arXiv by scientists in astrophysics and cosmology as well as the role of professional associations and learned societies in the publishing process for this community.

We will briefly explain the methods employed to compile this report. We will also briefly present the RIOJA toolkit before we try and identify costs in the publishing process associated with the functions of registration, certification, and awareness and archiving.

This report does not aim to provide a comprehensive report of actual journal publishing costings. Despite the fact that there are studies in existence that tried to document costs associated with journal publishing, the information presented there rarely corresponds to the actual costs of individual journal functions. In addition, the interviews with publishers and editors did not reveal any substantial information about costings that have not already been reported in the literature or are available on some publishers' websites. Where appropriate, this report aims to acknowledge studies conducted previously as pointers to further reading and, where applicable, to compare reported findings to observations made during the development and implementation of the RIOJA toolkit (described below).

We will conclude this report with some of the issues reported in the literature around sustainability of services and some brief suggestions for further work.

## *Definitions*

**Application Programming Interface (API)**
"*A language and message format used by an application program to communicate with the operating system or some other control program such as a database management system (DBMS) or communications protocol*".

(Definition by PC Magazine, Encyclopedia. Available at: http://www.pcmag.com/encyclopedia_term/0,2542,t=API&i=37856,00.asp)


**"Author pays" model**
A model of publication where the authors pays on acceptance of paper for publication and the paper then becomes available to the public in OA.


**Open access journal**
For the purposes of this report an open access journal is one which adheres to the principle of open access as this is defined in the JISC Open Access briefing paper version 2, see below. Available at: (http://www.jisc.ac.uk/publications/publications/pub_openaccess_v2.aspx)

*"The Open Access research literature is composed of free, online copies of peer-reviewed journal articles and conference papers as well as technical reports, theses and working papers. In most cases there are no licensing restrictions on their use by readers. They can therefore be used freely for research, teaching and other purposes".*


**Overlay journal**
For the purposes of this report an overlay journal is defined as a quality-assured journal whose content is deposited to and resides in one or more open access repositories.


**Overlay journal model**
An interactive model of scholarly publication between an open access repository (subject or institutional) and a scientific journal.

**Subject repository**
A repository that collects archives and disseminates works in a specific field of study. A well-known subject repository is arXiv (http://arXiv.org/).


**Institutional repository**
*"A university-based institutional repository is a set of services that a university offers to the members of its community for the management and dissemination of digital materials created by the institution and its community members. It is most essentially an organizational commitment to the stewardship of these digital materials, including long-term preservation where appropriate, as well as organization and access or distribution."*

Definition by Lynch, Clifford A. "Institutional Repositories: Essential Infrastructure for Scholarship in the Digital Age" ARL, no. 226 (February 2003): 1-7. http://www.arl.org/resources/pubs/br/br226/br226ir.shtml.

## *Background and statement of the problem*

Publishing research outcomes is important to the scientific community for various reasons. Cornet & Vollaard (2000) indicate that besides disseminating scientific knowledge, registration of a claim for new discovery and a "quality stamp" are the reasons why scientists publish. Besides social factors such as making research findings available and contributing to the advancement of knowledge, publishing is also a means for measuring quality of the work of scientists, allocate funding, and acknowledge contributions to knowledge. Therefore, it came as no surprise that 97% of the respondents in a recent RIOJA survey[1] indicated that papers for submission to peer review journals is the prime output of their research. Furthermore, scientists aim to publish in journals that are perceived as the best by their scientific community. Consequently, perceived quality of a journal is not only linked to the quality of the papers it publishes but also to its readership levels and reputation. The RIOJA project aimed to explore whether it is feasible to adhere to established quality standards using a model of publishing that interacts with open access repositories and, perhaps, could provide a cost effective solution to making research outcomes available to the public in a sustainable manner.

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, 2003) who discussed and compared functions of the existing publishing model what he referred to as the "deconstructed journal". Smith proposed *"...that all the roles of the academic journal could be fulfilled by a group of quasi-independent co-operating agents each playing a part in the activity of academic publishing. It also proposed that these agencies did not need to be organised or co-ordinated by a central publisher*". Similarly, Cornet & Vollaard (2000) suggested that more options for publishing scientific results should be given to the authors. They proposed a model of transferring copyright and the quality assessment of the paper to governmental organizations created for this purpose and acting as independent evaluators. Ginsparg (2003) took this further by arguing whether a more stringent evaluation could be more applicable to the current state of technological innovations and science publishing that peer review as it is currently conducted. Other models of peer review such as open and community peer review have also been documented (Rodriguez and colleagues, 2006; Casati and colleagues, 2007). In addition, 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.

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 report commission by JISC, 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

---

[1] Published results from the RIOJA project community surveys can be found at http://www.ucl.ac.uk/ls/rioja/dissem/

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 problem, in theory, to sustain 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 of providing 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, 2002; Crow & Goldstein, 2003a-c).

*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.

Journals are traditionally held to perform four "first order" functions (Meadows, 1974; Rowland [2002], Roosendaal and Geurts (1997) as cited by Prosser (2005)):
- 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

To those mentioned above Prosser adds the function of 'Reward' to the author.

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 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 resulting journal productions are not widely read when compared to the arXiv-held originals. During the development and implementation of the RIOJA tool (please 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.

*Learned Societies and scientific journals in Astrophysics and Cosmology*

Learned and Professional Societies have long been involved in the process of publishing scientific research. The American Astronomical Society (AAS) has been involved in the publishing and integration of publications to a wider information provision service including the Astrophysics Data Service (ADS), citation and management of datasets. Studies of the use and reading patterns of astronomers demonstrate the early adoption of electronic journals by astronomers and their close links with indexing services and the ADS (Tenopir et al, 2005). Although astronomers appear to use e-print services less compared to physicists, they were keener to adopt new electronic journal models and also tended to rely heavily on electronic communications. The use and impact that the ADS service had on the way the astrophysical community conducts research, as well as the impact that access to information via the ADS has contributed to the exploration of new bibliographic metrics for journal use and citation, have been reported in the literature (Kuntz et. al., 2000, 2005a, 200b,).

The Royal Astronomical Society is also involved in the publishing of scientific journals employing a commercial publisher (currently with Blackwell Publishing Ltd) maintaining though the scientific management of peer review as well as ensuring that the published papers adhere to ethical code of practice and promote the interests of the Society.

The Institute of Physics (http://www.iop.org) and the American Institute of Physics (http://www.aip.org) have both been pioneers in the publishing of scientific journals in the broader area of physics. Being a not-for-profit publisher denotes that any potential surplus and/or profit will be redirected to support the Society's aims and thus sustain if not increase benefits for its members.

Finally, the Association of Learned and Professional Society Publishers (http://www.alpsp.org/ngen_public/) has also been actively involved in researching and publishing issues around costs of journals publishing, methods of certification and peer review in particular and issues in scholarly communication in general.

## *Methodology*

This report utilizes a mixed methods research approach. In particular, it builds on the results from two community surveys (quantitative and qualitative) 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),
- Semi-structured face to face 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 explore whether an overlay journal model in astrophysics and cosmology could be viable in the long term.[2]

In addition, secondary research (desktop research) was employed to identify studies on the costs (alternative wording in the searches included economics and finances) of publishing scientific journals.

Detailed information on actual costings was not available to the researcher. This, allied to limitations of project scope means that this is not intended to be a comprehensive report on journal publishing costs. This report aims to acknowledge studies conducted previously as pointers to further reading and, where applicable, to compare reported findings against the development and implementation of the RIOJA toolkit described below.

---

[2] Results from the RIOJA project community surveys, both in the form of a) unpublished written report and b) peer reviewed book chapter and conference proceedings, can be found at http://www.ucl.ac.uk/ls/rioja/dissem/.

## *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 overlaying of academic journals on separate digital repositories. It is 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 so that they only see 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).

## Development

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

- Development of open API for communication between repositories and journals
- Development of  software for hosting overlaid journals using the API (outsourced to MetaOme in Bangalore)
- Demonstrate journal software using API implemented on arXiv.org repository Develop version of ePrints repository software to make complete open source package for any subject area (still in progress by MetaOme)

More information about the toolkit, short demonstration of the process of submitting a paper for publication using the RIOJA toolkit and access to the open source software can be found at the following URL : http://arxivjournal.org/rioja/. However, a typical workflow is presented below:

## Typical workflow

- Author submits paper to repository

- Repository optionally offers link to submit to a registered RIOJA-enabled journal, or author manually visits journal website

- Author provides repository ID to the journal (or this is provided by optional API)

- Journal extracts metadata from repository, displays summary to author, and confirms submission

- Journal checks repository paper status at regular intervals; once accepted by the repository the journal continues as below; if the paper instead becomes rejected or withdrawn by the repository the journal automatically rejects the paper.

- Optional API informs the repository that the paper is under consideration by the given journal (so the repository can report the status, and/or prevent submission to other journals)

- Journal proceeds with normal peer-review process: appointing editors, referees, etc, who write reports based on links to the paper on the repository

- Author may update paper on the repository in response to referee/editor feedback and journal re-submission can take place

- Journal finally either accepts or rejects the paper. If rejected optional API informs the repository.

- On paper acceptance the journal assigns a journal ID to the paper, lists it in its lists of accepted papers, optionally notifies the repository of the publication information, and makes metadata available by standard OAI interface

- Paper has page on the journal site giving metadata of published version, version-specific link to the published paper on the repository, and optional links to any future modified versions on the repository.

- In addition there are optional APIs allowing the journal to display trackback and download statistics for the paper

## *Costs of journal publishing*

The maturation of communication technologies and the economics of the journal publishing system over the last decades have been identified as two of the most influential factors that could transform the way scientific journal publishing is conducted (King & Tenopir, 2000, Cornet & Vollaard, 2000). In particular, technological innovation has already and could further improve the publishing system. This has also been confirmed in the RIOJA community surveys. In particular, interviews conducted with publishers of scientific journals indicated that the interest of academic and research staff in new publishing models is the prime driver for their adapting to technology challenges. For example, it was stated that one of the most successful journals of a publishing house, 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.

## 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 favorably disposed to the idea of overlay journal as a potential 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). Quality of other submitted papers is linked to the perceived quality of the journal and subsequently to its readership levels. The reputation of a journal, readership levels and impact factor are confirmed as important factors when scientists choose where to publish their research outcomes (Nicholas & Rowlands, 2005). Cornet & Vollaard (2000) argue that the first issues or initial year of running is the most crucial for any journal to establish reputation. Therefore, it is essential to enrol those factors that would contribute to the smooth running of the journal from the very start.

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 price per article processing varies by journal, discipline and usage. However, it was once more 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.

## Fixed and variable costs

Before we proceed to try and identify cost estimates for activities in each of the publishing functions it is useful to differentiate between 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 external factors and business models, e.g. in the case of a subscription based journal model, editorial costs are likely to remain the same (King, 2007; SQW Limited 2004) regardless of the number of subscribers. By contrast, variable costs refer to those that are subject to change, for example, again in a subscription based journal model, subscriptions maintenance costs could increase with the number of suscribers. Some of these costs could be higher when associated with particular business models (e.g. subscription based model versus "author pays" model) and publishing mediums (e.g. print versus electronic). It is possible that a cost recovery model such as "author pays" could be cheaper to sustain in an electronic environment, where variable costs are arguably negligible (Schonfeld and colleagues, 2004; Ginsparg, 2003). Similarly, an overlay journal model could perhaps be relatively cheap to sustain.

In the following section we will try to identify cost estimates associated with activities in each of the four functions in the publishing process: registration, certification, awareness and archiving. Areas in which an overlay model could contribute to reduced costs are highlighted.

## Registration

Some of the costs applying at the registration stage are submission costs. Costs at submission level include those for both accepted papers and rejected papers and are, in general, fixed costs. Furthermore, they include what is usually addressed as first copy costs. First copy costs are associated with article processing such as the work of the editor and editorial board (initial decision about paper's scope/suitability for publication, identification of reviewers and assignment of the review, 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 as well. Non-article processing or rejection costs incorporate some of the processes and associated costs as those of the 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. [3] Consultants in SQW Limited (2004) reported that first copy costs for a good to high quality journal are estimated around – average price - $1500 ($1650 including first copy and fixed costs). However, distribution costs do not vary with the number of subscriptions and in their majority most costs are fixed rather than variable. In studies by Garson and Ubell as cited by King (2007) fixed costs are estimated at 70-80% of the overall publishing process costs. Those studies included administrative, marketing and sales costs in those first copy attributions usually noted as indirect costs. Furthermore, it is even easier to separate and control submission costs if a submission fee and a publication fee is set separately. It is arguable that costs associated with the editorial process can not be reduced. However, it might be interesting to explore whether some of the editorial processes could be conducted in a quicker manner. For example, the initial paper

---

[3] King (2007) defines average prices as the division of the fixed, variable and indirect costs by the number of subscriptions that a particular journal comprises.

endorsement by the arXiv[4] for a submitted paper could speed up the initial editorial decision of whether a paper is within a journal's scope. Ginsparg (2001, 2003) estimates endorsement costs remain in the range of $1-$5 per submission, notably reported in the range of 100 to 10000 times lower compared to the existing peer review process. In addition, Harnad (2000) also indicates that conducting the peer review electronically and for papers residing in an open access archive could cost about 1/3 less of the actual page cost.


## Certification

Certification is the process where an independent party accredits and verifies conformity of a system/service/competence to specific standards so that quality is ensured. In the scholarly communication system peer review is one of the methods used to certify conformity to quality standards. The process of peer review has traditionally been conducted amongst peers of the same discipline. It has been applied to the assessment of books, allocation of grants and funding, scientific journal publishing. However, it should be noted that reviewers do not assess the way data was collected and analysed and in most occasions do not have access to primary data that led to the research outputs. Reviewers primarily judge whether a paper fits in with the journal's scope, whether it is well written (structurally and stylistically), whether the author(s) demonstrate knowledge of the field and acknowledge previous literature, whether scientific methods have been employed and adequately argued and also whether a paper reports on advancements/innovation in the field. Issues such as consistency, readability, interest to the journal's audience are also addressed. An example for the editorial criteria for a journal in this community, The Astrophysical Journal, can be found at the following URL: http://www.journals.uchicago.edu/page/apj/criteria.html

Peer review can be conducted in various manners. It can be blind, double-blind, or open. Blind peer review is when the identity of the reviewer is concealed to the author of the paper. Double blind peer review is when neither the author nor the reviewer knows whose the paper is or who is the person reviewing. Rowland [2002] describes the impact of technology in the way peer review is conducted and mentioned the evolution of the open peer review practice. Open peer review encompasses the concept of disclosing the identity of the reviewer to the author (on agreement) and the use of complementary methods facilitated by technology to support the process (such as voting, discussion in an electronic list, etc.).

In this community where use of electronic media is high, it could be expected that open peer review might be preferred. However, the RIOJA surveys' results showed that scientists believe that it is not always clear how the peer review is conducted and their comments suggested that, perhaps, there is room for improvement. Although no specific journals were researched, the comments the researchers made showed two schools of thought. The first called for a more open, publicly available peer review with the use of new technologies such as wikis, voting systems, being part of a discussion forum, etc. It was noted that the agreement of both the author and the referee would be essential in those situations.

---

[4] Endorsement in the arXiv is the process where scientists who have previously submitted their papers to the arXiv confirm that the submitted work of another author is within a specific subject area and of scientific value. Information can be found at: http://arxiv.org/help/endorsement

The second called for the maintenance of the anonymity in peer review, and for an exploration of the adaptation of more rigorous models of peer review which are applied in other disciplines. For example: "*One of the major flaws of current astronomy journals is that they use only a single referee. Compare this to other journals in physics, which can use up to 3 or 4 referees. Additional referees really help in the critique of a paper….".* Other models could also include adopting blind or double blind peer review. Educating the reviewers and having clear guidelines about what is expected from them was also commented upon. In detail: *"Besides, there is an appalling lack of appropriate referencing in all current journals. Somehow giving adequate credit to previous work carefully does not seem to be part of the educational training of astrophysicists".*

The interviews with members of publishing houses pointed to peer review as the most costly in the whole of the publishing process. Peer review, in terms of costs, is directly linked to the editorial costs. Authors do not get paid for writing a paper and referees are usually not paid for conducting the review. Editors and members of editorial boards in most occasions receive an honorarium that also covers any costs to their support such as administrative support. Indisputably, the manuscript submission system is central and vital in the management of communication between editors, referees and authors. King & Tenopir (2000) list the following activities in article processing: 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. There are also significant indirect costs - costs not directly associated with a particular process, such as administrative and managerial costs.

Rowland reports costs for peer review in the range of $200-$400 per paper, including administrative support, for a journal with rejection rate of 50%.

## Awareness

Awareness is directly associated with visibility. In this context and in the domain of astrophysics and cosmology services like the Astrophysics Data System and SPIRES are important. Therefore any scientific journal would need to ensure that papers that it publishes are made available via the ADS as well as through its own Web site.

arXiv already provides some awareness services with the alerting lists to which scientists can subscribe. Scientists may choose new (daily) or recent (weekly/monthly) submissions and they can also search for submitted papers by various elements (author, subject, title, journal identifier, comments, etc.). [5] The RIOJA survey confirmed the role that the arXiv plays in communicating and disseminating research in the fields of astrophysics and cosmology. In fact, about 80% of the respondents use the arXiv "new/recent" section to keep up to date with new research. In addition, when the scientists were asked "on finding an interesting title/abstract, where do you look for the full article", e-print repositories (such as the arXiv) were denoted as the first point of access for looking for the full text of a paper by 610 people (89%). Regarding searching for back literature, 68% of the scientists prefer the ADS service. "Other" responses pointed to the scientists' reliance on information gleaned from colleagues, journal alerting services, attendance at conferences and workshops, and visiting the SPIRES Web site.

---

[5] Search options for astrophysics can be found here: http://arxiv.org/find/astro-ph

In addition, RSS feed is now in operation (http://arxiv.org/help/rss) and several other awareness services are currently being explored such as track*backs (a method for Web authors to request notification when somebody links to one of their documents,* [6] http://arxiv.org/help/trackback*) and social bookmarking (the option for* users to store, *organize and manage bookma*rks of web pages, [7] http://arxiv.org/help/social_bookmarking).

The awareness functions provided by arXiv and other repositories could clearly reduce central overheads for a repository-overlaid journal.


## Archiving

Long term archiving and preservation of digital material have been identified as two of the most significant factors for the scientists that participated in the RIOJA online questionnaire survey. Furthermore, it was identified in that survey as one of the three most important functions in the process of science publishing in terms of how money should be apportioned. Pinfield (2007) emphasizes the role of preservation of digital content and its association with repositories. He suggests that models such as the one proposed by SHERPA DP where preservation is provided by a central agency for distributed repositories is a possible model for how preservation might be carried out in the future. Schonfeld and colleagues (2004) point out that if archiving is to be achieved, it must be paid for and argues about the role of libraries in this direction. Preservation of digital content is an area of vast changes and some initiatives that are already in operation include PORTICO (*http://www.portico.org/*), LOCKSS (Lots of Copies Keep Stuff Safe, http://www.lockss.org/lockss/Home) and PubMed Central (PMC, *http://www.pubmedcentral.nih.gov/*).

Overlay journals founded on top of repositories would be required to ensure sound arrangements for the long-term integrity and availability of accepted content, whether through journal-level or repository-level arrangements. arXiv is now operated and funded by Cornell University while it receives partial funding by National Science Foundation. It has a robust digital preservation policy, and some of the concerns raised by its community regarding supplementary material and file sizes have recently been addressed. [8] The long-term archive of any arXiv-overlay venture would be well supported.

In response to a survey question about factors that scientists regard as important when they consider publication, almost 30% noted that a print version of a journal is unimportant. In contrast to this finding, several respondents mentioned the role of print copies of a journal in maintaining an historical record and ensuring future access to the literature. This finding, perhaps, suggests the potential for collaboration between repositories and libraries and/or other memory organisations. Interestingly, a cost recovery option that some scientific journals employ include the on demand printing of past volumes for libraries or personal use. Such an option also serves to ensure that an archival copy of past issues resides in libraries. An example is the Journal of Machine Learning Research (http://jmlr.csail.mit.edu/) that uses Microtome Publishing

---

[6] Trackback at Wikipedia: http://en.wikipedia.org/wiki/Trackback
[7] Social bookmarking at Wikipedia: http://en.wikipedia.org/wiki/Social_bookmarking
[8] http://arxiv.org/new/#jun2008

(http://www.mtome.com/) for printing paper versions of the journal's issues at a cost of less than $250, including shipping and handling.

## *Sustainability*

Sustainable systems and services are a sought after outcome for those who fund their development, those that implement and manage them and also for their users. Zorich (2003) presents some of the requirements set by funding institutions that aim at achieving systems/services with sustainable outcomes. Although the focus of this study was the cultural heritage domain, such requirements could be applicable in other subject domains as well. Those include:

- Demonstration for demand for a particular system/service. It is suggested that a need for a particular system/service is what contributes to its long term use and therefore requirement to sustain. Tenopir and colleagues (2005) when they referred to the American Astronomical Society and the Society's involvement in the publishing of electronic journals they regarded it as a service designed to last: they have included several mirror sites, they produced the journal content in archival quality SGML paper and electronic versions are derived using software translation and they have demonstrated a commitment to maintain journals, by establishing a fund to pay for future conversions. Furthermore, as members of this community are active and consistent users of electronic resources in preference to print, the integration of electronic journals to a wider information provision system (ADS, citations, etc.) was a success factor to its use and sustainability.

- Development of a business plan. Swan (2007) discusses and suggests business models appropriate for open access systems/services. In particular, she recommends an approach that Clarke (2004) takes and forms a series of questions to answer in an appropriate business plan such as: who pays? pays what? for what? to whom? why? Such questions are meant to express the value of any proposed system/service, identify the relevant market and related segments, specify relations in the value chain, and specify how revenue will be generated. Reports commissioned by SPARC (Crow, 2002; Crow and Goldstein, 2003a-c) also discussed in detail different business models for both launching a new open access journal but also converting an existing journal to open access. They further discuss sources for cost –recovery and (author fees, grants, sponsorship, in-kind contributions, etc.)

- Engagement with scholarly opinion usually demonstrated by a significant contribution by academic staff or even leading project/development. Factors that would encourage scientists to publish in a scientific journal include, among other, the quality of other submitted papers, reputation of editorial board and peer review process (RIOJA community surveys, www.ucl.ac.uk/ls/rioja/project-docs/). Scientists tend to submit their papers to journals that they and their scientific community regard as the best in the field, also known as a journal's reputation. Reputation is built on readership levels and quality. Therefore, involvement of academic staff is crucial as a success factor.

- Institutional support (usually demonstration by cost sharing). In the case of a scientific journal publication examples that demonstrate the involvement of professional societies are again the American Astronomical Society and the Royal Astronomical Society. Examples of candidate cost-recovery business models include page charging, sponsorship, "value-added" charged services (such as the on-demand sale of bound volumes to libraries), and combinations of these solutions and others.

- Institutional infrastructure. Researchers at the OSS Watch (2007) discuss the different levels of sustainability particularly with relevance to open source software. They distinguish between levels of sustainability that conforms to meeting the immediate needs of the users where the expertise and skills for maintenance reside with the users and levels of sustainability that conforms to the needs of an entire

community and there is continuous development of both the software but also building programs for educating and training users.

- Specifications for long-term preservation.

## *Conclusions and suggestions*

The RIOJA project (http://www.ucl.ac.uk/ls/rioja) investigated the feasibility of an overlay journal model in collaboration with the arXiv and in the scientific domain of astrophysics and cosmology. The RIOJA toolkit comprises of a set of XML-based APIs for the exchange of data between digital repositories and journals to facilitate overlaying of academic journals on separate digital repositories. It has been 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. Therefore, we tried to identify estimates of costings associated with each of the journal publishing functions.

Despite the fact that there are studies in existence that tried to document costs associated with journal publishing, the information presented there rarely corresponds to the actual costs of individual journal functions. In addition, the interviews with publishers and editors did not reveal any substantial information about costings that have not already been reported in the literature or are available on some publishers' websites, e.g. PhysMath Central (http://www.biomedcentral.com/info/about/apcfaq). Interviewees suggested that the price per article processing varies by journal, discipline and usage.

The community surveys registered some encouragement for the overlay journal model in the fields of Astrophysics and Cosmology. However, general issues were raised about new and untested models of publishing, the overlay model included. It is clear that, for any new publishing model to succeed, it will have to address many 'traditional' publishing issues, among them impact, peer review quality and efficiency, building a readership and reputation, arrangements for copy-editing, visibility in indexing services, and long-term archiving. For a community such as astrophysics and cosmology in which publications benefit from the support of Learned Societies (such as the AAS and RAS) it is thought that perhaps it would be of interest to explore this further via case studies.

Suggestions for further work, in collaboration with a Learned Society, would be the following:

A) Run a trial study of the journal demonstrator (http://arxivjournal.org/) and solicit feedback from scientists about their preferred functionality.

B) Outline the publishing process costs in a newly launched overlay journal in collaboration with arXiv. For this purpose we would convert and extend the existing RIOJA demonstrator.

C) Outline the publishing process costs in the case of converting an existing journal to an overlay journal model in collaboration with arXiv. This would require experimentation with an already established and accredited scientific journal.

## References

Casati, F., Giunchiglia, F., and Marchese, M. 2007. *Publish and perish: why the current publication and review model is killing research and wasting your money*. Ubiquity 8, 3 (Jan. 2007), 1-1. DOI= http://doi.acm.org/10.1145/1226694.1226695

Clarke C (2004)  Open source software and open content as models for eBusiness. Available at: http://www.anu.edu/people/Roger.Clarke/EC/Bled04.html (Last accessed on 31/07/2008)

Cockerill, M. 2006, "Business models in open access publishing," in *Open Access: Key Strategic, Technical and Economic Aspects*, N. Jacobs, ed., Chandos Publishing, Oxford, pp. 89-95. Available at: http://demo.openrepository.com/demo/handle/2384/2367 (Last accessed on 31/07/2008)

Cornet, Maarten and Vollaard, Ben. 2000. Tackling the journal crisis. CPB Netherlands Bureau for Economic Policy Analysis, Van Stolkweg 14, 2585 JR, The Hague, The Netherlands. Available at: *http://www.cpb.nl/nl/pub/cpbreeksen/werkdoc/121/wd121.pdf* (Last accessed on 31/07/2008)

Content Complete Ltd 2007, *Business models trials: A report for JISC Collections and the Journals Working Group*. Available at: http://www.jisc-collections.ac.uk/media/documents/jisc_collections/business%20models%20trials%20report%20public%20version%207%206%2007.pdf (Last accessed on 31/07/2008)

Crow, R. 2002, *SPARC Institutional Repository Checklist & Resource Guide*, The Scholarly Publishing & Academic Resources Coalition (SPARC). Available at: http://www.arl.org/sparc/bm~doc/IR_Guide_&_Checklist_v1.pdf (Last accessed on 31/07/2008)

Crow, R. & Goldstein, H. 2003, *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 on 31/07/2008)

Crow, R. & Goldstein, H. 2003, *Guide to Business Planning for Launching a New Open Access Journal*, Open Society Institute. Available at: (Last accessed on 31/07/2008)

Crow, R. & Goldstein, H. 2003, *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 on 31/07/2008)

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 31/07/2008)

Ginsparg, P. (2003). Can Peer Review be better Focused? Available at: http://people.ccmr.cornell.edu/~ginsparg/blurb/pg02pr.html (Last accessed on 31/07/2008)

Ginsparg, P. (2001). Creating a global knowledge network. *Invited contribution, UNESCO Conference* Paris, 19-23 Feb 2001, Second Joint ICSU Press - UNESCO Expert Conference on Electronic Publishing in Science. Available at:

http://people.ccmr.cornell.edu/~ginsparg/blurb/pg01unesco.html (Last accessed on 31/07/2008)

Haggemann, M. (2006). SPARC Innovator: December 2006. Available at: http://www.arl.org/sparc/innovator/hagemann.html (Last accessed 31/07/2008)

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 31/07/2008)

Harnad, S. (2000). The invisible hand of peer review. Exploit, Issue 5. Available at: http://www.exploit-lib.org/issue5/peer-review/ (Last accessed on 31/07/2008)

King, D. & Tenopir, C. 2000. Scholarly Journal & Digital Database Pricing: Threat or Opportunity? At the Economics and Usage of Digital Library Collections Conference. Ann Arbor, Michigan, 2000. Available at: *http://www.si.umich.edu/PEAK-2000/king.pdf* (Last accessed on 31/07/2008)

King, D. (2007). The cost of journal publishing: a literature review and commentary. Learned Publishing, Vol. 20 (2), pp. 85-106.

Meadows, A.J. (1974) *Communication in Science*. London, UK: Butterworths.

Meadows, A.J. (1998) *Communicating Research*. San Diego, CA: Academic Press.

Michael J. Kurtz, Guenther Eichhorn, Alberto Accomazzi, Carolyn Grant, Markus Demleitner, Stephen S. Murray (2005a) Worldwide use and impact of the NASA Astrophysics Data System digital library. Journal of the American Society for Information Science and Technology, Vol. 56 (1), pp. 36-45. DOI: 10.1002/asi.20095. Preprint available at: http://cfa-www.harvard.edu/~kurtz/jasis-abstract.html (Last accessed 31/07/2008)

Michael J. Kurtz, Guenther Eichhorn, Alberto Accomazzi, Carolyn Grant, Markus Demleitner, Edwin Henneken and Stephen S. Murray, (2005b). The effect of use and access on citations, Information Processing & ManagementVolume 41 (6) , Special Issue on Infometrics, December 2005, Pages 1395-1402.( doi:10.1016/j.ipm.2005.03.010) Preprint available : http://arxiv.org/abs/cs/0503029v1 (Last accessed 31/07/2008)

M.J. Kurtz, G. Eichhorn, A. Accomazzi, C.S. Grant, S.S. Murray and J.M. Watson. (2000). The NASA Astrophysics Data System: Overview. Astronomy and Astrophysics. Supplement Series. Vol. 143 (1), pp. 41-59. Available at: http://aas.aanda.org/index.php?option=article&access=standard&Itemid=129&url=/articles/aas/pdf/2000/07/ds1780.pdf. Also, http://adsabs.harvard.edu/abs/2000A&AS..143...41K (Last accessed 31/07/2008)

Lawal, Ibironke (2001). Scholarly communication at the turn of a millennium: A bibliographic essay. Journal of Scholarly Publishing, April, pp. 137-154.

Lewis, A. 2007. RIOJA Journal-Repository APIs. Available at: http://cosmologist.info/xml/APIs.html (Last accessed 31/07/2008)

Nicholas, D. & Rowlands, I. 2005, *New journal publishing models: international survey of senior researchers*, Publishers Association and the International Association of STM Publishers. Available at: http://www.ucl.ac.uk/ciber/ciber_2005_survey_final.pdf (Last accessed on 31/07/2008)

Odlyzko, A (1997). The economics of electronic journals. First Monday, Issue 2 (8), .Available at: http://www.uic.edu/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/542/463 (Last accessed on 31/07/2008)

OSS Watch (2007). Sustainability study: a case study review of open source sustainability models. A JISC funded report. Available at: http://www.jisc.ac.uk/media/documents/programmes/distributedelearning/sustainabilitystudy-1%5B1%5D.0.pdf (Last accessed on 31/07/2008)

Pinfield, Stephen (2007). Can open access repositories and peer reviewed journals coexist? Serials 20 (3) 2007, pp.163-171

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 31/07/2008)

Rightcom Ltd 2005, *Business model for journal content: final report*, JISC. Available at: http://www.nesli2.ac.uk/JBM_o_20050401Final_report_redacted_for_publication.pdf (Last accessed on 31/07/2008)

Rodriguez, M. A., Bollen, J., and Van De Sompel, H. 2006. The convergence of digital libraries and the peer-review process. J. Inf. Sci. 32, 2 (Apr. 2006), 149-159. DOI= http://dx.doi.org/10.1177/0165551506062327. An arXiv preprint of this paper is available at:  arXiv:cs/0504084v3

Roosendaal, Hans E. and Peter A. Th. M. Geurts (1997). 'Forces and functions in scientific communication: an analysis of their interplay.' *Cooperative Research Information Systems in Physics*, August 31—September 4 1997, Oldenburg, Germany. http://www.physik.uni-oldenburg.de/conferences/crisp97/roosendaal.html

Rowland, Fytton [2002]. The Peer Review Process: A Report to the JISC Scholarly Communications Group Available at: *http://www.jisc.ac.uk/uploaded_documents/rowland.pdf* (Last accessed on 31/07/2008)

Schonfeld, Roger C., Donald W. King, Ann Okerson and Eileen Gifford Fenton (2004). The Nonsubscription Side of Periodicals: Changes in Library Operations and Costs between Print and Electronic Formats. Washington: Council on Library and Information Resources, 2004.

Smith, J W T. (1999). The deconstructed journal: a new model for academic publishing. Learned Publishing, Vol. 12 (2), pp. 79-91

Smith, J W T (2003). The Deconstructed Journal Revisited – a review of developments, in, ICCC/IFIP Electronic Publishing Conference - ElPub03 - From Information to Knowledge - held at Universidade do Minho, Guimarães, Portugal, 25-28 June 2003. Available at: *http://elpub.scix.net/data/works/att/0314.content.pdf* (Last accessed 31/07/2008)

SQW Limited (2004). Costs and business models in scientific research publishing: A report commissioned by the Wellcome Trust. Available at: http://www.sqw.co.uk/file_download/50 (Last accessed on 31/07/2008)

Suber, P. Open Access Overview: Focusing on open access to peer-reviewed research articles and their preprints. Available at: http://www.earlham.edu/~peters/fos/overview.htm (Last accesssed on 31/07/2008)

Suber, P. & Sutton, C. 2007. Society publishers with open access journals. SPARC Open Access Newsletter 115. 1-11-2007. Available at: http://www.arl.org/sparc/publications/society-publishers-with-o.html (Last accessed on 31/07/2008)

Swan, A. 2007, "The business of digital repositories," in  *A DRIVER's Guide to European Repositories*, K. Weenink, L. Waaijers, & K. van Godtsenhoven, eds., Amsterdam University Press, Amsterdam, pp. 15-48. Available at: http://dare.uva.nl/document/93898 (Last accessed on 31/07/2008)

Tenopir, Carol, Donald W. King, Peter Boyce, Matt Grayson and Keri-Lynn Paulson (2005). *Relying on Electronic Journals: Reading Patterns of Astronomers.* Journal of the American Society for Information Science and Technology, Vol. 56 (8), pp.786–802

Waltham, M. 2005, *Learned Society Open Access Business Models*, JISC. Available at: http://www.jisc.ac.uk/uploaded_documents/Learned%20Society%20Open%20Access%20Business%20Models.doc (Last accessed on 31/07/2008)

Zorich, Diane M (2003). A Survey of Digital Cultural Heritage Initiatives and Their Sustainability Concerns. Council on Library and Information Resources Available at:: http://www.clir.org/pubs/reports/pub118/contents.html (Last accessed 31/07/2008)