Assessing the Value of Open Access Information Systems:

Making a Case for Community-Based Sustainability Models

Oya Y. Rieger, Cornell University Library

ABSTRACT

In 2010, the Cornell University Library adopted a new business model to broaden the funding base for arXiv.org. Although this community-based interim model has garnered strong support, it also generated a range of questions in regard to the sustainability of open access online resources with a global user base such as arXiv. Making a case for such systems requires more than reliance on common quantitative measures that indicate submission and usage, statistics. Based on a sociotechnical systems approach, this article proposes a set of principles with which to evaluate return on investment as we consider the sustainability of open access subject repositories.

1. Introduction

Over the past two decades, open access digital repositories have become an increasingly vital component of the scholarly communication infrastructure. Such repositories are expected to facilitate broad and unrestricted availability of information and ensure enduring access to knowledge through preservation. Although these systems are technologically sound and sophisticated, questions about their significance and uptake remain to be addressed.[1] Several recent articles propose a set of performance measures with which to assess the value and usefulness of repositories. Almost all of them, however, are based on institutional repositories (Cassella, 2010; Carr & Brody, 2007; Thibodeau, 2007; Thomas, 2007; Swan & Brown, 2005). As Adamick and Reznik-Zellen (2010) argue, subject repositories are understudied in the scholarly communication and information science literature. In the meantime, subject

---

[1] Currently, several open access repositories are in use; they offer rich sets of functionalities to facilitate depositing, accessing, and retrieving scholarly materials. Rieh et al.'s (2007) census of institutional repositories (IRs) reveals, however, considerable uncertainty regarding the underlying practices, policies, and content. As Foster and Gibbons (2005) point out, without content an IR is just a set of empty shelves. Thus far, IRs have only partially fulfilled their original intent, and their adoption by potential users has been limited. The quantity of deposited content remains quite modest. Although the present situation can be traced in part to structural issues such as copyright and tenure publication requirements, most impediments to widespread implementation relate directly to a gap between developers' and potential users' perceptions of the scholarly communication landscape (Rieger, 2008). Recent studies stress the importance of working with faculty closely to better understand their work habits and research contexts. An excellent example of this approach is provided by Maness et al. (2008) in their creation of personas to describe various classes of potential IR users to facilitate increased participation.

repositories such as arXiv are often regarded as successful scholarly communication forums, especially in comparison with institutional repositories.

Using the arXiv sustainability initiative as a case study, this article suggests an assessment metrics for evaluating the significance and utility of subject repositories. The recent economic downturn has created even greater pressure for online repositories to demonstrate their return on investment (ROI). Relying on such measures is particularly important when we are making decisions on how to invest our limited institutional budgets in supporting specific scholarly communication initiatives.

1.1 The arXiv Sustainability Initiative

Since its launch in 1991, arXiv has achieved iconic status as an effective online distribution system and is often cited to illustrate digital repositories' potential role in transforming scholarly communication. One of the premises of arXiv has been making science more democratic by allowing for the rapid worldwide dissemination of scientific findings. Figure 1 illustrates the international reach of arXiv. Since it moved to Cornell in 2001, the Cornell University Library has provided the bulk of arXiv's operating costs, which are projected to be approximately $500,000 for 2011.[2]

In January 2010, Cornell announced an interim business plan for generating funds through recurring subsidies from libraries at academic institutions, research centers, government laboratories, and other organizations that are the heaviest users of arXiv. The library has established a voluntary institutional contribution model and invited pledges from the top 200 institutions accounting for more than 75 percent of annual institutional downloads. The model entails a tiered structure of annual support requests that is similar to those of many other open-access funding models. The three-tiered institutional support model is based on the previous calendar year's download activity, suggesting institutional contributions within a range of $4,000 to $2,300 per year. Cornell University Library will continue to provide 15% of arXiv's operating budget, an amount many times greater than the support requested from other heavy user institutions. The repository will remain free to readers and submitters.

---

[2] The arXiv budget for 2011 is available at http://arxiv.org/help/support/2010_budget. It is based on an estimate and will be updated throughout the year to reflect the actual expenses.
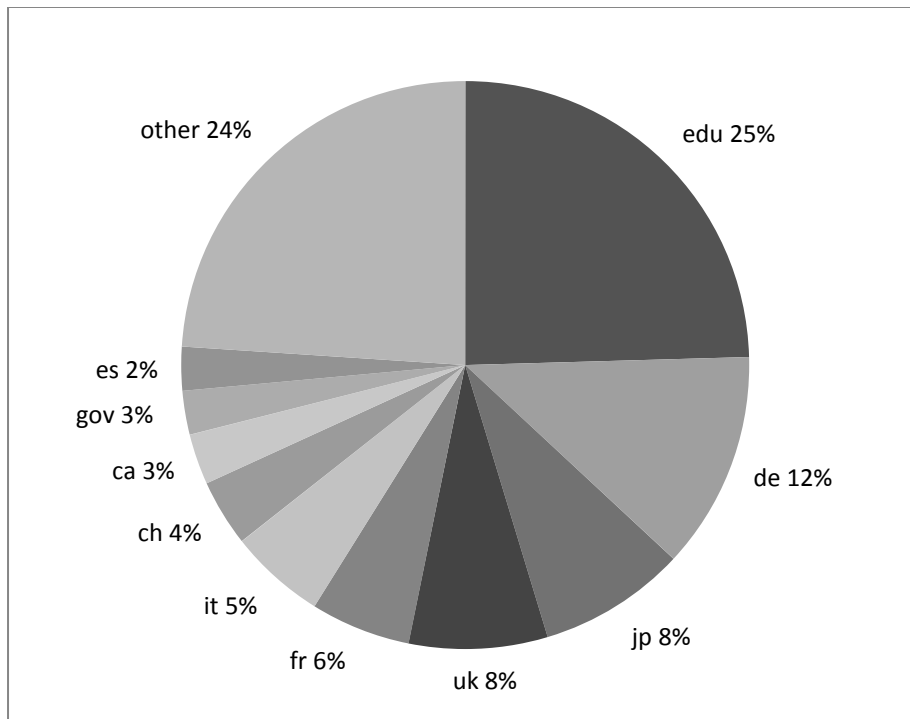
Figure 1: arXiv institutional downloads at main site by Internet domain of institutions (2010)

The arXiv sustainability initiative that began in June 2009 was necessitated by significant budget cuts that required the Library to review its programs and funding sources. Although financial pressures triggered the planning process, the Library was already engaged in investigating alternative funding sources to support and further develop the archive. For instance, the funding strategies under consideration had included establishing an endowment and considering grant opportunities. The first phase of the sustainability planning process included a landscape analysis and a survey of arXiv stakeholders' positions and opinions about arXiv's future. Also critical during this assessment phase was expanding our understanding of income models for open access and understanding the pros and cons of emerging practices such as those presented by Crow (2009). The Library decided to implement an interim business model for 2010–2012 pending completion of a thorough (thus time-consuming) business planning process that would involve many stakeholders including scientists, libraries, research centers, societies, publishers, and funding agencies.[3]

Since the announcement of the interim collaborative business model, as of December 2010, Cornell University Library secured pledges from 125 institutions, totaling $340,000 in contributions. The Library has enjoyed broad international support from Australia, Canada,

---

[3] The arXiv white paper at http://arXiv.org further describes the business planning process and addresses questions raised by stakeholders during the input-gathering process.

China, France, Germany, India, Israel, Japan, Switzerland, the United Kingdom, and the United States. The upshot thus far validates the interim model and suggests that this approach is likely to be a viable component of a long-term business strategy. Although the interim institutional contribution model has garnered strong support, we realize that, as a transitional strategy, it needs to be further assessed and developed.

1.2 Sustainability of Open Access Repositories

The sustainability report prepared by Guthrie, Griffiths, and Maron (2008) provides a comprehensive review of a variety of business models for supporting online academic resources. The authors define sustainability as "the ability to generate or gain access to the resources financial or otherwise needed to protect and increase the value of the content or service for those who use it." Therefore, keeping open access academic resources (such as arXiv) sustainable entails not only covering the associated operating costs but also continuing to enhance their value based on the needs of the full range of user communities. Furthermore, sustainability involves running a robust technical operation and addressing the digital preservation requirements to ensure the repository's longevity and usability. Such a financial commitment is likely to exceed a single institution's resources. Since scholars worldwide depend on the stable operation and continued development of arXiv, sustainability is best assured by aligning revenue sources with constituents who realize value from their use of arXiv, and by reducing dependence on Cornell University Library's budget. One of the goals of the business planning initiative has been to engage the institutions that benefit from arXiv to assist in defining the future of the service.

Open access is a scholarly communication model and should not be conflated with free production. The Budapest Open Access Initiative (2002) characterizes the outcomes of scholarly communication process as public goods and advocates broad, completely free, and unrestricted online access free of financial, legal, or technical barriers.[4] The research and library community's reaction to escalating serials prices played a significant role in ushering in the open access movement. Recognizing the potential of open access for removing the *price and permission barriers* that may undermine library efforts to provide access to scholarly communication process outputs, the research and library community has been a vocal and active advocate of open access. Nevertheless, as Crow (2009) reminds us, it is important to distinguish the philosophical and social perspectives on open access from its economic aspects. The compelling argument behind the open access model is that it improves the equity, efficiency, and reach of scholarly communication. Open access does not entirely remove 'fees' and 'access limitations' but it replaces and reconfigures them for the key stakeholders in the scholarly communication endeavor.

Traditional publishing is sustainable because the research and academic community supports such enterprises through subscriptions that are charged against acquisitions budgets. As SCOAP3 illustrates, such funding sources can also support open access, but using them in this way requires a new way of thinking and planning (Gentil-Beccot & Schimmer, 2008). Repositories, whether they are institutional or subject-based, entail various expenses and require clearly

---

[4] Budapest Open Access Initiative: http://www.soros.org/openaccess/read.shtml

defined financial responsibility and accountability. One of the premises of open access ideology is that the overall costs of providing open access to the scholarly literature are far lower than are the costs of traditional forms of dissemination. Formal publishing involves a range of expenses including acquisition, administration of the peer review process, copyediting, marketing, and order fulfillment. The advent of Web-based publishing has eliminated 'printing and distribution' as one of the cost elements; however, running electronic publishing systems can be equally or even more expensive. Although they are likely to lower the overall cost of dissemination, successful open access initiatives necessitate new cost-recovery models and financing mechanisms. The few extant cost-analysis case studies do not provide sufficient data for comparing conventional with open access publishing budget models in order to confirm the hypothesis that open access will significantly lower costs.

As the Cornell team surveys administrators, managers, and scientists from libraries and research centers about the future of arXiv, one frequently asked question addresses arXiv's value proposition. In other words, "What is in it for me and why should my organization provide funds for arXiv?" Another common question has been whether the current business model will induce other institutions that manage globally used open access online academic resources to establish similar collaborative sustainability strategies. The crucial question is, "Does this business model represent a new trend and, if so, how do we set priorities and decide which systems to support with our finite institutional funds and competing priorities?" It is likely that community-based financial arrangements will serve as one of the revenue streams needed to sustain open access information systems. As illustrated by the financial models of Wikipedia and the Stanford Encyclopedia of Philosophy, we view this business trend as inevitable—a necessary step in supporting open access to knowledge. Consequently, we would do well to answer the following question: "How do we assess the value of a digital repository in order to justify and encourage investment in sustaining such resources?"

Common ROI indicators such as usage and content growth patterns are useful in making a case for supporting online academic resources. As we seek broader institutional support, however, we need to expand beyond such common metrics and establish a set of criteria that will help us assess the value and stability of a repository. I argue that it is a mistake to reduce sustainability issues to economics. As we seek to institute appropriate metrics for evaluating the sustainability of such repositories, we need to view them as sociotechnical systems in acknowledging that they are embedded in a rich ecology of scholarly practices and norms.

3. Repositories as Sociotechnical Systems

The concept of a sociotechnical system involves the interrelatedness of *social* and *technical* aspects of an information system. Repositories are not merely information technologies; they exist as networks of content, people, practices, and policies. arXiv as a sociotechnical system integrates material aspects of information technologies with the associated social and work practices. The sociotechnical approach emphasizes an information system's interactions with elements of its broader context including institutions, knowledge, society, and knowledge-production regimes (Van House, 2003; Lievrouw & Livingstone, 2006; Bijker, 1995). Therefore, arXiv consists of technical systems and standards, activities and practices involved in developing and using the system, and the social arrangements and organizations that provide it with a structural framework.

In the early days of repository development efforts, research on and accounts of digital repositories privileged material aspects such as system design, metadata standards, interoperability protocols, and technical requirements. The past decade has witnessed growing interest in evaluating repository technologies and services from a usability standpoint (Ferreira et al., 2008). Although there is increasing emphasis on sociocultural perspectives, only a handful of studies examine repository technology from such a holistic perspective. As Williams and Lawton (2005) point out, much of the research and literature continues to focus on the perceived benefits of repositories as a new scholarly communication technology, such as expanding access to scholarly information, reducing the costs of producing and acquiring scholarly publications, and gaining control of the scholarly communication process from commercial publishers. There is also an extensive body of literature on technical aspects of digital repositories in the information, computer, and library science fields addressing issues such as system architecture, interoperability, metadata, and digital preservation (Knowles, 2010; Biswas & Dibyendu, 2010; Hitchcock, 2007). However, little attention has been devoted to understanding the negotiation process surrounding repository development and implementation efforts from the perspective of the shifting interests of relevant stakeholders.

At the foundation of the explanatory framework of sociotechnical systems is the concept of relevant social groups (Bijker, 2005) that share a particular set of meanings about a technology or an information system. In the case of arXiv, the relevant social groups include a wide range of stakeholders such as researchers, publishers, societies, libraries, and open access advocates, to name just a few. These groups interpret such an online communication forum variously based on their needs, roles, goals, values, and motivations. These stakeholders also vary in their ability to influence the development, application, and acceptance of a repository (Rieger, 2008; Rieger, 2007).

The following section offers a set of assessment criteria based on the characterization of arXiv as a sociotechnical system, implying that it comprises a complex system of technologies, people, practices, norms, and values that are related in emergent and unpredictable ways. Such a perspective takes into consideration the ecology of scholarly communication and factors in the well-established and embedded infrastructures of traditional scholarly communication models.

4. Assessing the Value of Open Access Information Systems: Criteria for Evaluating Return on Investment

As Cornell announced its interim business model, the repository's usage statistics and repository collection patterns were highlighted as indicators of a successful online system.[5] As of December 2010, arXiv includes over 645,000 e-prints that are used by hundreds of thousands of researchers from all over the world. arXiv has transformed the scholarly communication infrastructure of multiple fields within the broader physics discipline and plays an increasingly prominent role in mathematics, computer science, and other disciplines (Ginsparg, 2007). Based on a budget of $380,000 and 30 million paper downloads during 2009, each e-print costs merely 1.3 cents. Based on 60,000 new submissions during 2009, the cost per submission is less than seven dollars per e-print. These important indicators are relevant to any effort to secure support; however, we

---

[5] Cassella (2010) offers a practical metrics for user-based performance indicators for institutional repositories. Several these measures also apply to subject repositories.

need to consider additional measures that factor in a repository's sociotechnical aspects. The following discussion proposes six principles that can be used in assessing the value of an open access system. It offers examples based on arXiv; however, it is important to note that this particular service does not necessarily fulfill all these requirements. The current interim planning stage has been undertaken not only to investigate how to diversify revenue models but also to ensure that arXiv strives to meet the proposed principles and can set an example of transparency and reliability as a community-supported service.

4.1 Deep Integration into the Scholarly Community and Scholarly Processes

As I was explaining our efforts to raise external funds for supporting arXiv at a Physics Department meeting, one faculty member asked, "Isn't arXiv the most important thing Cornell University Library is doing?" This perspective illustrates the essential success principle for arXiv—it has become firmly entrenched in scientific research workflows and its value and role in supporting scholarly communication is widely acknowledged. Through Paul Ginsparg's leadership, which is rooted in both the academic and information science communities, the service has consistently focused on the disciplinary cultures represented in the digital repository and on community needs (Ginsparg, 2001). arXiv is the primary daily information source for hundreds of thousands of researchers in physics, and plays an increasingly prominent role in mathematics, computer science, and other related fields. During 2010, submissions in the math category reached 24 percent of the total arXiv submissions, exceeding the high-energy physics subject category, which has constituted the core discipline of the repository.

Disciplinary characteristics, work practices, and conventions of academia play an important role in researchers' assessment and appropriation of information communication technologies. Repository deployment cannot be fully understood without comprehending how a specific technology is embedded in its social context. The information and communication technology integration characteristic of disciplinary communities often mirror underlying differences in epistemic cultures. Recent studies that have explored evolving scholarly communication patterns in various disciplines report that the views and practices of researchers diverge sharply from strategies promoted by information providers and policy makers (Harley et al., 2010; Research Information Network & the British Library, 2009; Velden & Lagoze, 2009). These studies identify a gap between the actual practices and perceptions of researchers and the vision that motivates advocates of open access, data sharing, preprint servers, and scholarly blogs in transforming scholarly communication. Harley et al. (2010) caution that enthusiasm for the development and adoption of information technologies should not be conflated with the hard reality of highly competitive and complex professional environments.

Several studies of arXiv explore how the disciplinary characteristics of high-energy physics influenced the successful appropriation of arXiv within the scholarly community (Merz, 2006; Fry, 2006; Gunnarsdóttir, 2005; Bohlin, 2004; Kling, 2004; Cronin, 2003; Kling & McKim, 2000). For instance, reliance on pre-prints among physicists is a cultural practice that precedes the development of arXiv—efforts to collect unpublished research reports in high-energy physics began in 1962 (Merz, 2006; Bohlin, 2004; Kling, 2004). Existing scholarly publishing patterns and intellectual property policies often reveal the readiness of a discipline to share published or unpublished papers online. Other indicators include task dependency and competition. Physics, especially in the high-energy domain, represents a close-knit culture in which physicists work

together on joint projects (Fry, 2006). Because physicists work in a competitive fashion as well, they need to expedite the publishing process to be able to claim ownership of or take credit for reported results and achieve immediate visibility (Fry, 2006; Merz, 2006). Their work is cumulative, so physicists rely heavily on robust sharing of results via arXiv. High-energy physics is a highly visible field and active researchers are keenly aware of each other's ongoing projects so physicists have considerable confidence in the robustness of collective arrangements for validating new knowledge (Bohlin, 2004; Cronin, 2003).

arXiv is closely associated with high-energy physics because that discipline constituted the core content of the e-print server. As illustrated in Figure 2, arXiv submissions by subject reflect a range of topics that is not limited to high-energy physics. Figure 3 demonstrates the shifting balance of subject coverage, as math and computer science contents are the fastest-growing subject categories. arXiv user studies have predominantly focused on how the physics community interacts with the system. One of Cornell's goals during the business planning process is to expand our understanding of the use patterns and needs of the disciplines that increasingly rely on arXiv to facilitate their day-to-day interactions.
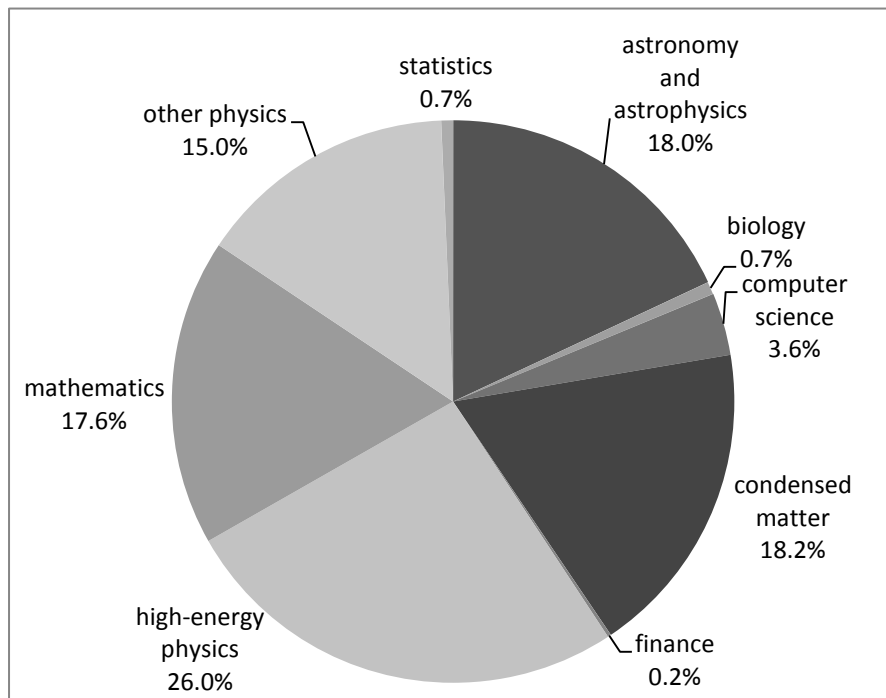

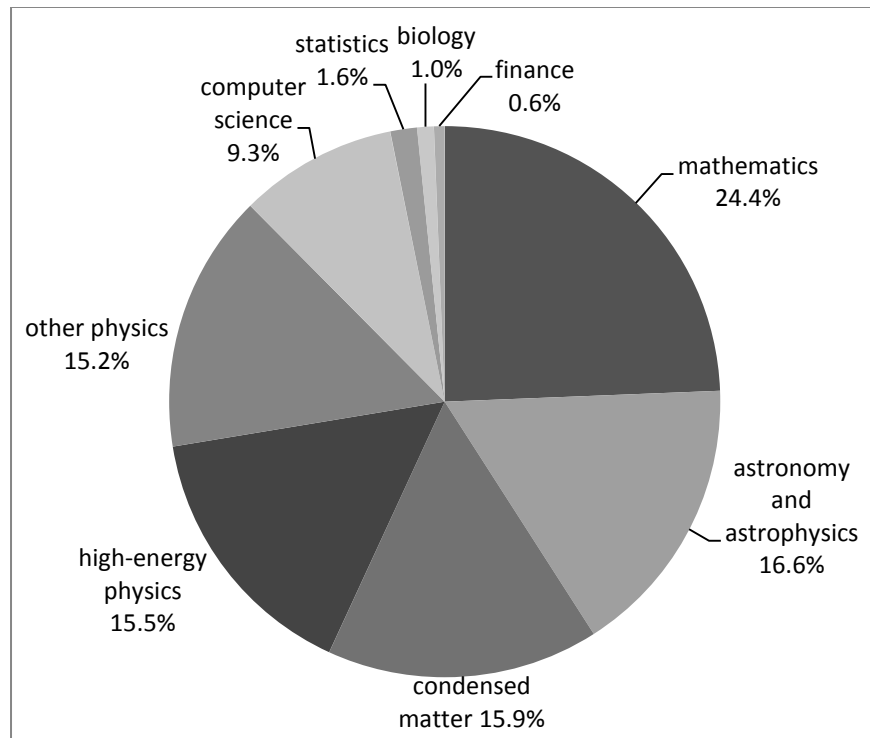**Figure 2: arXiv submissions by subject, 1991-2009**

Figure 3: arXiv submission by subject, 2010

4.2 A Clearly Defined Mandate and Governance Structure

Although best practices in developing technical architectures and associated processes and policies underpin a trusted digital repository, organizational attributes are equally important. The Trustworthy Repositories Audit & Certification: Criteria and Checklist (TRAC) tool emphasizes that organizational attributes affect the performance, accountability, and sustainability of repositories.[6] The first criteria in the TRAC assessment tool are governance and organizational viability. Similarly, subject repositories must have clearly defined mandates and associated governance structures to reflect a commitment to the long-term stewardship of a service.

Although we perceive the exploration and planning process as a collaborative endeavor, Cornell University Library is ultimately responsible for carrying out the policies and decisions as well as for managing the overall arXiv operation throughout its business planning activities. This is a critical principle that must be observed in order to maintain a stable and reliable service point. The arXiv Website clearly states the mandate and governance of the service:

> Started in August 1991, arXiv.org (formerly xxx.lanl.gov) is a highly-automated electronic archive and distribution server for research articles. Covered areas include physics, mathematics, computer science, nonlinear sciences, quantitative biology and statistics. arXiv is maintained and operated by the Cornell University Library with

---

[6] The Trustworthy Repositories Audit & Certification: Criteria and Checklist (TRAC) is available at http://www.crl.edu/sites/default/files/attachments/pages/trac_0.pdf

guidance from the arXiv Scientific Advisory Board, arXiv Sustainability Advisory Group, and the help of numerous subject moderators.

Cornell University Library relies on advice from two groups in managing day-to-day operations as well as planning the future of arXiv. The arXiv Scientific Advisory Board is a group of international scientists and researchers who serve as consultants to Cornell University Library regarding the operation, technology, policies, and future direction of arXiv. Cornell University Library looks to the board members for input and guidance on decisions of strategic importance. The physics, mathematics, computer science, quantitative biology, quantitative finance, and statistics archives each have advisory committees, which are overseen loosely by the scientific advisory board.

The arXiv Sustainability Advisory Group assists the Library in evaluating opportunities and risks associated with alternative financial and technological strategies required to maintain arXiv as a reliable and innovative service. They provide feedback in setting a time line and strategies for transitioning from the current support model that is based solely on institutional contributions to a diversified, longer-term business model. Leveraging existing professional connections and venues, this group acts as a liaison to the broader community by representing arXiv in related forums and gathering input that is required for business planning purposes. Its members also contribute to the communication and promotion of decisions and strategies to key stakeholders (users, libraries, publishers, research institutions, etc.) regarding the arXiv business model.

Organizational mandates are instrumental in defining sustainability strategies. Based on a review of available funding models and an extensive survey of arXiv stakeholders, the sustainability team considered several support options that are compatible with the repository's mandate. We are committed to maintaining arXiv as an open-access resource that anyone may use to download and read articles as well as to allowing free submissions so that all appropriate articles can be accepted. In order to attract a broad range of international contributions to the repository free of financial expectations on the part of authors, the team decided not to pursue contributions from users who deposit or access materials. This decision was primarily guided by the repository's mandate and mission.

4.3 Systematic Development of Content Policies

Content curation and stewardship roles have traditionally been shared by libraries and publishers—publishers with a focus on creation and distribution and libraries specializing in discovery and preservation. In digital information environments, repository services such as arXiv blur the distinction between libraries and publishers. An essential criterion in assessing open access online resources is the availability of clearly defined collection policies and submission guidelines that reflect the content curation and stewardship role of the hosting institution. To that end, the content policies for arXiv are included on the arXiv webpage:

> arXiv is an openly accessible, moderated repository for scholarly articles in specific scientific disciplines. Material submitted to arXiv is expected to be of interest, relevance, and value to those disciplines. arXiv reserves the right to reject or reclassify any submission. Submissions are reviewed by expert moderators to verify that they are topical and refereeable scientific contributions that follow accepted standards of scholarly

communication (as exemplified by conventional journal articles). As an electronic archive, arXiv makes a commitment to provide persistent access to all announced submissions. arXiv is thus maintained with a focus on the perpetual availability of submissions. This is accomplished in part by controlling the types of files that can be uploaded to arXiv, as well as restricting changes that can be made after submissions are announced. In order to preserve the scholarly record, submissions are not removed from arXiv after they have been announced.

Although arXiv is not peer-reviewed, submissions are reviewed by subject-based moderators. Additionally, an '*endorsement*' system is in place to ensure that content is relevant to current research in the specified disciplines.[7] It is also critical to have clearly articulated policies about the copyright status of the deposited materials as well as conflict management processes (such as responding to concerns in regard to rejected submissions). arXiv supplements the traditional publication system by providing immediate dissemination and open access to scholarly articles (which often appear later in conventional journals). The arXiv help pages clearly state that the repository is neither a venue for refereeing nor a distribution mechanism for work that is considered unpublishable. The moderation process is essential to ensuring that submissions are of value to arXiv communities, but there is also a limit to the ability of administrators and moderators to provide feedback on submissions that are determined to be inappropriate for arXiv.

Cornell regularly fields requests to extend arXiv to include other subject areas. Adding a new subject area cannot be taken lightly and such addition involves a complex technical and social process: exploring the user-base and use characteristics pertaining to the subject area, establishing the necessary advisory committees, and recruiting moderators. Due to its limited resources, the Library has adopted a measured approach to expansion because there is significant organizational and administrative effort required both to create and to maintain new subject areas.

arXiv complements, rather than competes with, the commercial and scholarly society journal publishing market. Among the challenges are ensuring the authority and integrity of e-prints and distinguishing between succeeding versions, such as a pre-print paper and its published version in a scholarly journal. It is therefore critical that the repository and preservation community address the versioning of scholarly articles, tracking them from initial submission to pre-print archive to final publication in a formal scholarly journal. Other challenges include linking the burgeoning corpus of institutional repositories with related subject repositories in order to achieve version control as well as creating a critical mass of related materials on particular topics. There is a great potential for subject and institutional repositories to function in a complementary fashion by leveraging their particular strengths (Darby et al., 2008). This would represent yet another measurement of success. There are standards, technologies, practices, and policies that would allow such synergies; however, there is need for a broad architectural map to conceptualize such an information environment.

4.4 Technology Platform Stability and Innovation

---

[7] arXiv's submission and moderation guidelines are described at http://arxiv.org/help/primer.

Keeping open access academic resources sustainable involves not only covering the operational costs but also continuing to enhance their value based on the needs of the user community. Such a goal requires a stable but flexible IT architecture that supports incremental modifications and allows for innovation and enhancements. There is often an inherent tension between technological stability and innovation. The repository needs to be fully operational to fulfill its daily production functions, processes, and procedures. Having a dependable system in place is critical in order to provide reliable access to services on which end-users can depend. Although stability and consistency are important service attributes, also essential is keeping pace with evolving user needs through research and development (R&D) projects. Accordingly, assessing resource needs and planning for investments in staffing and equipment are challenging tasks on both the maintenance and R&D fronts. However, given the uncertainties associated with the development and testing of new features and services, an innovation agenda needs to be carefully thought out in order to ensure that operational stability is not undermined. Ideally, there should be complementary streams of resources to support both operational and research activities; nevertheless, the hard financial realities require that we carefully balance these two critical program areas.

The existing repository ecology is a complex of architectures and features that are optimized to fulfill the specific needs of institutional, subject, or archival repositories. The landscape is becoming even more heterogeneous with the addition of scientific social networking sites that profile local scholarly activities and open data initiatives that focus on data curation models. As we evaluate a specific repository, it is essential to consider this rich context and understand how the service fits within the broader framework. In particular, we need to factor in the following three aspects:

- Interoperability arrangements that link a given repository to related systems, services, and communities;

- Features that support supplementary information objects such as underlying data, auxiliary multimedia content, and research methodologies;

- Functionality and arrangements that lower barriers to contributing content to multiple complementary repositories.

As we witness the proliferation of information sources (such as pre-prints, published versions, supplementary information), the discovery and access domain is also becoming increasingly complex. An additional factor in this rich and complex repository ecology is the primacy of search engines such as Google in discovering and accessing a wide range of information resources.

Among the critical roles of repositories is facilitating the preservation function. Digital preservation (used interchangeably with 'archiving') refers to a range of managed activities to support the long-term maintenance of bitstreams, thereby ensuring that digital objects are

usable.[8] However, ensuring enduring access involves more than bitstream preservation. It must provide continued access to digital content through various delivery methods. According to the *Preservation Management of Digital Material Handbook*, 'preserving access' entails ensuring the usability of a digital resource by retaining all quantities of authenticity, accuracy, and functionality deemed to be essential. As we assess the value of subject repositories, it is important to differentiate between bitstream preservation and preserving access.[9] Also, some subject repositories may opt to institute best practices for managing digital content but may not be in a position to assume full preservation responsibilities. It is therefore critical to assess the clearly defined roles and associated procedures that pertain to this responsibility.

The arXiv software has been developed in-house over many years and this has both benefits and burdens associated with it. While the underlying technology for arXiv has been updated throughout its history, arXiv requires significant internal re-engineering to support a shifting technological landscape, increased growth and use, and to ensure the sustainability of the service (Ginsparg, 2007). As the content expands beyond high energy physics, the service needs to be able to accommodate the publication and communication habits of recently joined communities.

Cornell University Library is in the process of assessing repository technologies to replace the existing system that supports the access, browse, and search processes. Creating a generalized architecture will facilitate efficient technology management processes. To maintain its role as an innovator in scholarly communication, it is essential that arXiv supports the deposit and archiving of supplementary information objects that are associated with a given paper. This functionality can be achieved through ingesting supporting digital objects within the repository (a feature that already exists) or through establishing robust connections with specialized service providers. The priority during the next couple of years will be to create a sustainable information technology architecture that will support the daily processes associated with submission, ingest, access, and search processes. In the meanwhile, the arXiv team will leverage related projects to pilot data repository services through collaborations, such as the one with the Data Conservancy initiative at Johns Hopkins University.[10]

The sustainability of arXiv also depends on enabling interoperability and creating efficiencies among repositories with related and complementary content to reduce duplication of efforts. Organizations with institutional repositories are usually keen to have them used for a range of needs, and would like to avoid the need for authors to make multiple deposits. SWORD (Simple Web-service Offering Repository Deposit) aims to lower the barriers to contributing content to

---

[8] The Trusted Digital Repositories report defines *digital preservation* as "the managed activities necessary for ensuring both the long-term maintenance of a bitsteam and continued accessibility of content." *Trusted Digital Repositories: Attributes and Responsibilities. An RLG-OCLC Report*. May 2002. Available at http://www.rlg.org/legacy/longterm/repositories.pdf.

[9] *Bitstream preservation* aims to keep digital objects intact and readable. It ensures bitstream integrity by monitoring for corruption of data fixity and authenticity and protecting digital content from undocumented alteration, securing the data from unauthorized use, and providing media stability. *Digital objects* are items stored in a digital repository and in their simplest form consist of data, metadata, and an identifier.

[10] Funded by a National Science Foundation (NSF) grant, the Johns Hopkins University Sheridan Libraries is involved in a project that aims to build a data research infrastructure for the management of the ever-increasing amounts of digital information created for teaching and research.

multiple repositories.[11] arXiv implemented the SWORD protocol for automated deposit over a year ago. This protocol enables both multiple deposits from a single tool and deposits from another repository. Although it has not yet been used to address the 'multiple deposit problem,' it has been successfully used by journals and conference systems depositing in arXiv.

A key item on arXiv's technical agenda is clearly identifying the preservation responsibilities of Cornell University Library (Rieger & Warner, 2010). The Library has an extensive technical infrastructure and associated best practices for maintaining its digital collections and repositories. Currently, arXiv is maintained as a component of this common technology service framework, including a nascent Fedora-based archival repository and an associated preservation policy framework. In addition, there are multiple mirror sites through which the continuity of arXiv service can be maintained.[12] One outstanding issue continues to be determining whether the preservation service needs to be expanded beyond a bitstream preservation strategy to a full-fledged enduring-access model that requires certification. A critical issue underlying this question is understanding the relationship of arXiv to third-party preservation services such as Portico, LOCKSS, and CLOCKSS. arXiv not only includes pre-prints but also contains a growing body of final publications that might already be preserved through other archiving services. The arXiv team aims to bring clarification to these questions as a component of the current sustainability initiative.

4.5 Reliance on Business Planning Strategies

Business models are created in order to link an organization's mission to a plan of sustainable modes and models (Bishof & Allen, 2004). The process involves the articulation of a mission, values, organizational structures, and financial plans. Business plans offer an overall view of a given product, relevant user segments, key stakeholders, communication channels, competencies, resources, networks, collaborations, cost structures, and revenue models (Osterwalder, Pigneur, & Tucci, 2005). Such plans represent an institution's underlying strategies and choices for creating and capturing value within a network of stakeholders.

The primary purpose of a business planning process is conveying a clear value proposition to justify investing in a service or product by its potential users. The value proposition describes the benefits that a product or service provides. In other words, value propositions respond to question: Why should an institution purchase your product or service? Since the focus of the value proposition is on the customer, it should be stated from the end-users' perspective. Value propositions may be based on a range of characteristics such as service features, customer support, product customization, and economical pricing. The key challenge in creating a value proposition is addressing the needs of all stakeholders. For instance, in the case of arXiv, the stakeholders include scientists, libraries, research centers, societies, publishers, and funding

---

[11] SWORD is a lightweight protocol for depositing content from one digital repository to another in order to lower the barriers to depositing at multiple sites. Additional information is available at http://swordapp.org/.

[12] Simeon Warner's October 26, 2010 posting on the liblicense-l mailing list responds to a question about what would happen in the unlikely event that Cornell did not continue to operate arXiv by describing the safety network formed through the establishment of mirror sites and also through other independently operated services such as eprintweb.org with content overlap (http://www.library.yale.edu/~llicense/ListArchives/1010/msg00073.html).

agencies. Although they are likely to have common goals, each group attaches value to a specific aspect of arXiv. For instance, from the end-users' perspective, scientists' highest priority for arXiv is likely to be the robustness and reliability of the repository and access features.

Business models also convey financial plans. In a collaborative business model such as the interim arXiv strategy, it is critical to clearly define and justify the pricing model and the budget to understand how revenues are being generated and spent. Table 1 includes the projected arXiv budget for 2011. Maintaining, supporting, and further developing a repository involve a range of expenses including management, programming, system administration, curation, storage, hardware, facilities (space, furniture, networking, phone), research and training (such as attending meetings and conferences), outreach and promotion, user documentation, and administrative support. The hosting institution must be willing to continue investing resources in operating the system (such as Cornell's commitment for covering 15% of arXiv's expenses in perpetuity in addition to some indirect costs). Also, the governing organization needs to have a viable and culturally sensitive exit strategy if a hosting institution should find it necessary or advantageous to withdraw its support.

Table 1: Projected 2011 arXiv budget.

Cost Category		Cost
Personnel	**Staffing**	
User support	2.5 FTE + 0.36 FTE student	
Programming and system maintenance	2 FTE	
Administrative and business model support	0.15 FTE	
Management	0.5 FTE	
Subtotal for personnel (including benefits)		$411,796
Personnel (in-kind)		
Management in Cornell Library and in Information Science	0.3 FTE	
Associated research effort in Information Science	>1 FTE	
arXiv moderation	130+ moderators, varying time commitments	
Subtotal for personnel (in-kind)		$0
Non-personnel		
Server amortization, hosting, hardware maintenance, storage and backup		$47,666
Network bandwidth and telephony		$15,000
Staff computers, software and supplies		$1,000
Staff conference/meeting attendance and arXiv sustainability advisory group travel support		$20,700
Preservation services		$8,000
TOTAL ESTIMATED COSTS FOR CY 2011		*$504,162*

The business model needs to be responsive to the shifting ecology of scholarly publishing and factor in the positions and initiatives of key stakeholders. As noted above, arXiv complements, rather than competes with, the commercial and scholarly society journal publishing market. Based on requests from several publishers and societies with publications in physics and mathematics, the arXiv team has been exploring strategies for expanding the current institutional contribution model to include input from relevant publishers and societies. Also included in the business planning process is looking into other potential funding sources such as related foundations and agencies.

Finally, it is also critical to consider the role of the Sponsoring Consortium for Open Access Publishing in Particle Physics (SCOAP3) initiative in the financial planning process (Douglas, 2009; Gentil-Beccot & Schimmer, 2008)). arXiv would potentially be a beneficiary of redirected

funding administered by the SCOAP3 consortium. It is not clear, however, when this initiative will meet its annual funding goal. Also it is important to note that because the SCOAP3 initiative is restricted to high energy physics and particle physics content only (which represents between 18% and 40% of submissions to arXiv depending how broadly the subject area is construed), if it is successful it could potentially subvent a similar fraction of arXiv's operating costs. The arXiv team continues to monitor the development of SCOAP3 and its impact on our long-term plans.

The current sustainability model represents a short-term strategy for the next three years. Over the next few years we will develop a durable business plan that provides a strategic framework within which to protect and increase the value of arXiv for those who use it. Ideally this will comprise a blend of ongoing underwriting from Cornell University Library and support from the academic library community and research centers. It might also include support from scholarly societies, an endowment, or funding agencies.

4.6 Implementation of User-Based Strategies and Feedback Cycles

As user communities (depositors and users of content) expand, the institutions that host repositories must continue to modify and mainstream support and make sure that there is an efficient user support system in place. It is necessary to institute mechanisms by which to gather, evaluate, and respond in a timely fashion to feedback from subscribers, supporters, submitters, users, and other interested entities regarding policies, features, and system performance. Although questions and requests received from users form an important channel of feedback, it is also important to systematically observe distinct user communities in order to understand their scholarly communication patterns and factor in their evolving user needs and requests.

Systematic gathering of information about users and their usage patterns can be instrumental in balancing the power and potential of information technologies with the appropriate needs and workflows. Although it is tempting to add new and innovative features, we must strike a balance between system adjustments and evidence-driven improvements based on actual user input on the one hand and the addition of experimental and novel functionality that introduces new modes and options on the other hand. Advisory and steering groups also act as sources of feedback and input as well as assisting the home institution in setting priorities.

5. Looking Ahead

The arXiv case study presented in this article illustrates the need to assess repositories holistically by taking into consideration a range of lifecycle and usability issues as well as factoring in changing patterns and modes characteristic of scholarly communication. As the library and research community collectively addresses the creation and management of community-based infrastructures, we need to factor in financial needs, usability factors, innovation in discovery and access, and enduring access. This article proposed six assessment principles based on a sociotechnical approach:

1. Clearly defined mandate and governance structure

2. Deep integration into the scholarly community and scholarly processes

3. Systematic development of content policies

4. Technology platform stability and innovation

5. Reliance on business planning strategies

6. Implementation of user-based strategies and feedback cycles

By analyzing sociocultural factors we can attain a better understanding of how repository technologies should be designed and implemented, and improve communication and promotional activities to encourage their appropriation. Information and communication technologies need to be designed, assessed, and promoted with an appreciation that they are a joint product of technical features and intergroup negotiations. To design effective information and communication technologies, we need a better understanding of the associations among the information practices, institutions, and the social and material foundation of the scholarly communication processes (Rieger, 2008).

arXiv has been at the forefront of the open access movement and has served as a model for many other initiatives. As open access publishing initiatives proliferate and users rely more heavily on free resources, it is inevitable that educational and cultural institutions will need to collaborate in experimenting with new funding strategies. What is essential is that organizations pursuing such undertakings share their experiences and lessons learned with the broader community in order to collectively enhance our understanding of issues and the pros and cons of potential strategies. To this end, Cornell University Library is committed to continue discussing the sustainability planning process and outcomes with our colleagues through multiple forums and channels.

REFERENCES

arXiv Business Model White Paper. January 2010. http://arxiv.org/help/support/whitepaper

Adamick, J., & Reznik-Zellen, R. (2010). Trends in large-scale subject repositories. *D-Lib Magazine* 16(11/12)

Bijker, W. E. (1995). *Of bicycles,bakelites,and bulbs: toward a theory of sociotechnical change*. Cambridge, MA: MIT Press.

Bishof f, L. & Allen, N. (2004). *Business planning for cultural heritage institutions*. Council on Library and Information Services. http://www.clir.org/pubs/reports/pub124/contents.html

Biswas, G. & Dibyendu, P. (2010). An evaluative study on the open source digital library softwares for institutional repository: Special reference to Dspace and greenstone digital library. *International Journal of Library and Information Science*, 2(1), 1-10. http://www.academicjournals.org/ijlis/PDF/pdf2010/Feb/Biswas%20and%20Paul.pdf

Bohlin, I. (2004). Communication regimes in competition: The current transition in scholarly communication seen through the lens of the sociology of technology. *Social Studies of Science*, 34(3), 365-391.

Carr, L. and T. Brody (2007). Size isn't everything: sustainable repositories as evidenced by sustainable deposit profiles. *D-Lib Magazine*, 13(7/8). http://www.dlib.org/dlib/july07/carr/07carr.html

Cassella, M. (2010). *Institutional repositories: An internal* and *external perspective on the value of IRs* for *researchers' communities. LIBER Quarterly*, 20 (20) http://liber.library.uu.nl/publish/articles/000503/article.pdf

Cronin, B. (2003). Scholarly communication and epistemic cultures. *New Review of Academic Librarianship*, 9(1), 1-24.

Crow, R. (2009). *Income models for open access: An overview of current practice*. Washington, D.C.: Scholarly Publishing & Academic Resources Coalition http://www.arl.org/sparc/bm~doc/incomemodels_v1.pdf

Darby, R. M., Jones, C. M., Gilbert, L. D., & Lambert, S. C. (2008). Increasing the productivity of interactions between subject and institutional repositories. *New Review of Information Networking*, 14(2), 117-135.

Douglas, K. (2009). Exploring the SCOAP3 model for high energy physics : A new innovation in open access. *College and Research Libraries*, 70 (6), 348-350.

Ferreira, M., Baptista, A.A., Rodrigues, E., & Saraiva, R. (2008). Carrots and sticks: Some ideas on how to create a successful institutional repository. *D-Lib Magazine*, 12(1/2). http://www.dlib.org/dlib/january08/ferreira/01ferreira.html

Foster, N.F., & Gibbons, S. (2005). Studying faculty to improve content recruitment for institutional repositories. *D-Lib Magazine*, 11(1). http://www.dlib.org/dlib/january05/foster/01foster.html

Fry, J. (2006). Coordination and control across scientific fields: Implications for a differentiated e-science. In C. Hine (Ed.), *New infrastructures for knowledge production: Understanding e-science* (pp. 167-187). Idea Group Inc.: Hershey, PA.

Gentil-Beccot, A. & Schimmer, R. (2008). Libraries can make open access happen today by simply redirecting subscription funds. *Liber Quarterly*. 18 (3/4) 449-458. http://liber.library.uu.nl/publish/articles/000274/article.pdf

Ginsparg, P. (2007). Next-generation implications of open access. *CT Watch Quarterly*, 3(3), http://www.ctwatch.org/quarterly/articles/2007/08/next-generation-implications-of-open-access/

Ginsparg, P. (2006). As we may read. *Journal of Neuroscience, 26(38)*, 9606-9608, www.jneurosci.org/cgi/content/full/26/38/9606

Ginsparg, P. (2001). Creating a global knowledge network. In R. Elliot & D. Shaw (Eds.) *Proceedings of the second ICSU Press-UNESCO Conference on Electronic Publishing in Science*. http://arXiv.org/blurb/pg01unesco.html

Gunnarsdóttir, K. (2005). Scientific journal publications: On the role of electronic preprint exchange in the distribution of scientific literature. *Social Studies of Science*, 35(4), 549-579.

Guthrie, K., Griffiths, R., & Maron, N. (2008). *Sustainability and revenue models for online academic resources. An Ithaka report.* http://www.ithaka.org/ithaka-s-r/strategy/sustainability-and-revenue-models-for-online-academic-resources

Harley, D., Acord, S.K., Earl-Novell, S., Lawrence, S., & King, C.J. (2010). *Assessing the future landscape of scholarly communication: An exploration of faculty values and needs in seven disciplines*. UC Berkeley: Center for Studies in Higher Education. http://cshe.berkeley.edu/research/scholarlycommunication/

Hitchcock, S., Brody, T., Hey, J. & Carr, L. (2007) Digital preservation service provider models for institutional repositories: Towards distributed services. *D-Lib Magazine*, 13 (5/6). http://www.dlib.org/dlib/may07/hitchcock/05hitchcock.html

Kling, R. (2004). The Internet and unrefereed scholarly publishing. In B. Cronin (Ed.), *Annual review of information science and technology* (pp. 591-637). Medford, NJ: Information Today.

Kling, R., & McKim, G. (2000). Not just a matter of time: Field differences and the shaping of electronic media in supporting scientific communication. *Journal of the American Society for Information Science and Technology,* 51(14), 1306-1320.

Knowles, J. (2010). Metadata in repositories: An overview. *Welsh Repository Network*. http://www.wrn.aber.ac.uk/objects/metadata_overview/

Lievrouw, L.A. & Livingstone, S. (2006). Introduction to the updated student edition. In L.A. Lievrouw and S. Livingstone (editors). *Handbook of new media: Social shaping and consequences of ICTs (*pp. 1–14). London: Sage.

Maness, J.M., Miaskiewicz, T., & Sumner. T. (2008). Using Personas to Understand the Needs and Goals of Institutional Repository Users. *D-Lib Magazine*. 14/9-10. http://dlib.org/dlib/september08/maness/09maness.html

Merz, M. (2006). Embedding digital infrastructure in epistemic culture. In C. Hine (Ed.), *New infrastructures for knowledge production* (pp. 99-119). Hershey: Idea Group Inc.

Osterwalder, A., Pigneur, Y., & Tucci, C. L. (2005). Clarifying business models: Origins, present, and future of the concept. *Communications of the Association for Information Systems*, 16(1). Article 1.

Research Information Network, & the British Library (2009). *Patterns of information use and exchange: case studies of researchers in the life sciences*. http://www.rin.ac.uk/our-work/using-and-accessing-information-resources/patterns-information-use-and-exchange-case-studie

Rieh, S.Y, Markey, K., Beth St. J., Yakel, E, & Kim, J. (2007). Census of institutional repositories in the U.S.: A comparison across institutions at different stages of IR development. *D-Lib Magazine*, 13(11/12). http://www.dlib.org/dlib/november07/rieh/11rieh.html

Rieger, O.Y. (2008). Opening Up Institutional Repositories: Social Construction of Innovation in Scholarly Communication. Juornal of Electronic Publishing, 11(3). http://dx.doi.org/10.3998/3336451.0011.301

Rieger, O.Y. (2007). Select for success: Key principles in assessing repository models. *D-Lib Magazine*, 13(7/8), http://www.dlib.org/dlib/july07/rieger/07rieger.html.

Rieger, O.Y. & Warner, S. (2010). Sustainability Case Study: Exploring Community-Based Business Models for arXiv. *International Conference on Preservation of Digital Objects*. Vienna, Austria.

Swan, A. & Brown, S. (2005). *Open Access self-archiving: an author study*. Technical Report , External Collaborators, Key Perspectives Inc. http://eprints.ecs.soton.ac.uk/10999/

Thibodeau, K. (2007). If you build it will it fly? Criteria for success in a digital repository. *Journal of Digital Information*, 8(2). http://journals.tdl.org/jodi/article/view/197/174

Thomas, G. (2007). Evaluating the impact of the institutional repository, or positioning innovation between a rock and a hard place. *New Review of Information Networking*, 12(2), 133–146.

Van House, N., Bishop, A. & Buttenfield, B. (2003). Digital libraries as sociotechnical systems. In A. Bishop, N. van House, & B. Buttenfield (eds.), *Digital library use: Social practice in design and evaluation* (pp. 1-21). Cambridge, MA: The MIT Press,

Velden, T., & Lagoze, C. (2009). Commentary: communicating chemistry. *Nature Chemistry*, 1, 673-678.

Williams, S. & Lawton, F.D. (2005). eScholarship as socio-technical change: theory, practice and praxis. *Third International Evidence Based Librarianship Conference*. Brisbane, Austria. http://conferences.alia.org.au/ebl2005/Williams.pdf