11 | Repository as Publishing Platform

Simone Sacchi and Mark Newton

Within academic libraries, programs around digital repositories and scholarly publishing have matured in tandem over the first part of the 21st century. Under the programmatic umbrella of *scholarly communication,* libraries have employed staff to work on common digital platforms to support institutional aims for partnering in the creation of and access to scholarly materials originating with authors, editors, and other content producers at their home institutions. Across the platforms that enable these programs and the library staff acting as agents to operate them, there are many correlations. In some instances, it is precisely the same staff members and the same platforms performing the core functions of both the repository and scholarly publishing programs. This chapter examines the functions and processes across both of these areas of programmatic emphases, making a more precise specification of this correlation. As repository- and library-based publishing programs are shown to share essential components, some conclusions about the appropriateness for integrating these programs, as well as for communicating the publishing role of the repository and the implication for libraries, are drawn out for discussion.

PRELIMINARY DEFINITIONS

The following discussion necessitates some definitional boundaries around *repository* and *publishing* for context.

Repository: By *repository,* we mean *institutional repository* (or IR), which is network-connected infrastructure that supports the discovery, access,

and preservation of research materials produced by the faculty, staff, and students of individual institutions of higher education. Repositories, as discussed here, are library-administered programs, and local collection policies for content acquisition may vary. We distinguish here between *mediated repository* and *nonmediated repository*.

- **Mediated repository:** By *mediated repository* we mean a repository where the content submitted goes through a process of review and refinement in its description typically conducted by professional librarians and other library staff before acceptance.
- **Nonmediated repository:** By *nonmediated repository* we mean a repository where publication after submission in expedited after little or no human processing. *Nonmediated repositories* also typically enable the submitting user to make changes in the content of the repository, including changes in the files and related metadata description.

Publishing: By *publishing* here we restrict the context to *online scholarly publishing* (or e-publishing), the process of selecting, reviewing, refining, compiling, and making available the results of research and scholarship (such as articles into a peer-reviewed online journal).

Publication: By *publication,* however, we discuss the abstraction of communicating the results of science and scholarship, which may be accomplished through repositories and journals, among others.

Stewardship: By *stewardship* we intend *digital stewardship*, the series of managed activities to ensure access to digital content into the future and through changes in technology.

INTEGRATING REPOSITORY AND PUBLISHING PROGRAMS: A RATIONALE

There are many available examples of integration of digital repository and publishing programs in academic libraries. At the staffing level, it is often the role of a single person, small cluster, or FTE fraction to accommodate the functions of both programs, as is evident in the latest job advertisements seeking library professionals to staff scholarly communication programs (Bonn, 2014).

It is also true that platform investments commonly accommodate both *publishing* and *repository* functions. A recent survey of the respondents to a call for information on publishing activity in academic libraries

suggests exactly this: 41% of respondents report using Digital Commons (a hosted hybrid journal publishing/repository solution). An additional 29% of respondents support publishing activity through the DSpace repository platform,[1] and yet another 15% do so on the Fedora Commons repository platform (Lippincott, 2014). Such crossover is hardly surprising, given the publication role repositories fulfill for institutions and the limited resources that libraries can allocate to development and areas of growth.

Despite the prevalent use of repository software among library publishers, the intentional separation of repository and publishing programs is also apparent at the platform level. The majority of respondents (mirroring results from a series of surveys over the past decade)[3] use the open source Open Journal Systems[2] to provide local editors with a manuscript solicitation, review, and publication toolkit. Even from an infrastructure perspective, sharing the same *platform* does not necessarily mean *service* integration: content in university-published journals is not always available in the repository, and similar processes (such as submission, review, archival, and dissemination) may be implemented separately.

Notwithstanding this apparent integration, repository and journal publishing programs may be administered separately, each with its own agenda, goals, and means. Library publishing programs indicate the intent of academic libraries to participate in the creation of new knowledge,[4] while repositories may be understood as vehicles for the distribution of scholarly communication and not, as Clifford Lynch notes in his landmark paper, as a "call for a new scholarly publishing role for universities" (Lynch, 2003). Reticence to formally, publicly affiliate repository and publishing programs may still be observed. Surveys of publishing activity in libraries routinely ask respondents to segregate repository and publishing activity in an apparent attempt to capture discrete pockets of activity. Open access advocates may find the publishing function of repositories to be an unwelcome conflation as well, diverting scarce resources and diluting the core message to potential content depositors.

Preliminary Observations

A call for an integrated approach between repositories and university-published journals is not new within the scholarly communication community. Soon after the publication of the Open Archives Initiatives[5] Protocol for Metadata Harvesting (OAI–PMH), advocates for a change in scholarly

communication envisioned a global adoption of the OAI Protocol such that "overlay journals"[6]—that is, journals implemented and managed as service providers over content in a repository—could take advantage of a distributed network of interoperable repositories sharing their content. Their ambition, however, has yet to fully materialize.

The authors' proposal for an integrated model here is somewhat similar in approach, at least functionally, but it is applied in the context of repository and library publishing programs within an institution. This perspective is driven by an analysis of their internal situation at Columbia University where repository and publishing programs coexist at the Center for Digital Research and Scholarship (CDRS).[7] Whether repository and publishing programs are already established enterprises within an institution or just at a preliminary analysis stage, library administrators of such programs might benefit from the analysis presented here and the emerging assessment framework.

The Columbia University Case Study

The Center for Digital Research and Scholarship (CDRS) at Columbia University Libraries (CUL)[8] is engaged in both a mature repository program, with its Academic Commons[9] research repository, and a thriving journal publishing program, with more than 20 publishing partners across the university. Although the collective efforts of the center have always been driven by mutually fruitful conversations between the staff responsible for both repository and journals publishing, the two programs have been developed in parallel since the center's inception in 2007.

Repository Program: Academic Commons runs on a Fedora Commons–based infrastructure (hereafter: *Fedora*[10]). The Fedora repository instance is shared with other digital collection projects within the CUL system. The Academic Commons collection, however, is independently indexed and presented online through a faceted-browse search-and-discovery front-end. Custom applications (e.g., self-deposit interface, cataloging tool) have been developed to manage mediated ingest and quality control over the object metadata descriptions.

Library Publishing Program: Journal publishing at CDRS is achieved in a variety of context-dependent ways. Partner projects vary by platform (e.g., Open Journal Systems and WordPress), by type of content

published (full articles, abstracts only, supplementary content affiliated with the journal brand), and by build approach (collaborative development or CDRS-managed). The approach to partnership development (and a loose adherence to prescribed project tiers) therefore coheres the program above all else. Much of the team's recent work has focused on the development of custom journal publication templates to expedite production and improve the prospects for scaling to accommodate additional partners (Newton, Cunningham, & Morris, 2013; Perry, Borchert, Deliyannides, Kosavic, & Kennison, 2011).

To this point, integration between the two programs has been managed through specific terms that permit repository contribution of journal content as specified in the Master Service Agreements outlining the primary responsibilities of the partners (i.e., the editors). Center staff working on the repository and journals communicate the specific parameters using issue-tracking software, and additions of CDRS-published journal content to the repository are committed manually by repository staff.

A significantly tighter platform integration between the programs, however, has been proposed. Advantages could then be realized at several levels (from the practical and administrative to the programmatic and strategic):

- Reducing the overall number of platforms managed within the center, thus improving prospects for allocating limited development staff to work within a more aligned and sustainable codebase, thus scaling up the number of partner projects to meet demand
- Taking advantage of the preservation functionality of the repository infrastructure and avoiding content duplication
- Multichannel dissemination, facilitating discovery, reach, and impact of the submitted content from different interfaces
- Repurposing of content and metadata from a unique authoritative source, improving consistent dissemination and interoperability capabilities
- Coordinating outreach opportunities: leveraging both the repository and publishing program user bases for coordinated messaging and outreach

Integration, however, presents new challenges. From a technical perspective, platform-level integration means purposeful segue from well-worn

tools and approaches to ones less familiar. Also, WordPress and Open Journal Systems employ one set of technologies and languages, while the applications developed to manage content within Fedora use another, making the transition or alignment less straightforward. Further, integration reveals swaths of policy questions to be resolved:

- All content published in Academic Commons is *necessarily* freely accessible, but not all of CDRS' partners produce open access journals. This is not a problem today as CDRS does not facilitate limited or gated access to journal articles through its partnerships. Still, the policies of the repository will constrain the range of possibilities for individual editorial policies in the matters of persistence, access, and reuse. It is unclear whether program integration would necessitate a series of policy reconciliation discussions.
- The matter of persistent identifier assignment is already complex. Persistent DOIs are created for published repository content, regardless of whether the files themselves are exact copies for which the original publisher also created an identifier. Identifiers are also prepared for a number of CDRS partner journals. Reigning in the multiplicity of identifiers at play as well as the locations and contents of their resolution will be necessary to further align the programs.
- It is presumed further that program integration will apply first to prospective partnerships and published content. How then to retrospectively reconcile the bodies of published content? To date, content published through CDRS partnerships has duplicative access points, retrievable both on the original publishing platform and the repository.

All of these concerns can, of course, be managed practically. Despite caveats and complications, the authors believe the benefits of deliberate program integration exceed them.

FUNCTIONS AND PROCESSES IN SCHOLARLY COMMUNICATION: AN ANALYSIS OF INTEGRATION STRATEGIES

The approach presented here is based on mapping the elemental functions in scholarly communication against processes in repository and library publishing programs to identify and assess integration strategies. The emerging

framework—presented in the next section—is based on the analysis of the Columbia case study, but may be generalizable to other institutional contexts as an analytical device for assessing the feasibility and appropriateness of similar integration efforts.

A Functional Perspective on Scholarly Communication

Roosendaal and Geurts in an influential paper (Roosendaal & Geurts, 1997) presented an analysis of scholarly communication in terms of core functions—*Registration*, *Certification*, *Awareness*, and *Archiving*—that can be summarized as follows.[11]

> **Registration** allows claims of precedence for a scholarly finding.
> **Certification** establishes the validity of a registered scholarly claim.
> **Awareness** allows actors in the scholarly system to remain.
> **Archiving** preserves the scholarly record over time.

A 2002 position paper prepared for SPARC by Raym Crow (Crow, 2002) compares, with respect to these functions, the traditional academic journal system model of scholarly communication to a new online disaggregated model. The analysis demonstrates how the elemental scholarly communication functions, many of which are already performed (if not organized) by members of academic institutions, can be directly and effectively enabled and sustained within the institutions themselves. This visionary approach relied on the aforementioned distributed global network of interoperable repositories sharing their content via the OAI-PMH. While institutional repositories have constantly grown both in numbers and in content, the conditions—in terms of collective effort and shift in the academic culture and practice—required to realize such an interoperable infrastructure never really obtained.

The functions of scholarly communication are therefore covered in an environment where the traditional journal publishing system coexists with institutional repositories (see Figure 11.1).

Intuitively, the *Registration* and *Awareness* functions are fulfilled by both the traditional journal publishing system and institutional repositories: they both capture and record attribution and date of submission, and both provide means to the scientific community to access the submitted content

✔: enabled o : incomplete	Registration	Awareness	Certification	Archiving
Traditional publishing system	✔	✔	✔	o
Repository system	✔	✔	o	✔

Figure 11.1. Scholarly communication functions enabled by the traditional publishing system and repository programs.

(once accepted in their final version). The other two functions—*Certification* and *Archiving*—when present, are typically expressed differently within the traditional publishing system and institutional repositories.

Repositories are not typically equipped to adequately fulfill the *Certification* function: the credibility granted by the "associative certification" applied by a recognized academic institution to content within its repository is insufficient to certify content quality. The peer-review process traditionally associated with journal publication, alternatively, persists as a widely acceptable means of certifying the quality of research within disciplinary communities, and publishing in established peer-reviewed journals is still a major component of the promotion and tenure system in academia.

Journal publishers operating in a traditional publishing environment used to rely on academic libraries for the *Archiving* function over print content. Although joint initiatives between participating libraries and traditional publishers have been developed to solve archiving and preservation issues over publisher-licensed digital content (e.g., LOCKSS,[12] CLOCKSS,[13] and Portico[14]), individual academic institutions retain an archiving interest over the entire range of scholarly outputs produced by their communities. Institutional repositories play an active role in this context, enabling the *Archiving* function within academic institutions by adopting platforms (e.g., Fedora) with which to manage digital content and support auditing functions such as those required by the ISO 16363/TDR Trusted Digital Repository.[15]

The SPARC paper imagines the outgrowth of repository programs to happen amidst a scholarly communication landscape where journal production is managed primarily by commercial and scholarly society stakeholders. Further, it does not explicitly address the presence of journal

Function	Process	Actor	Sponsor	Program
Registration	Submit to the repository	Authors	Academic Institution	*Repository Program*
	Submit to the journal	Repository team	Journal	*Library Publishing Program*
Certification	Associative certification	Institution	Academic Institution	*Repository Program*
	Peer review	Referees	Journal	*Library Publishing Program*
Awareness	Repository access / API	Repository Program	Academic Institution	*Repository Program*
	Journal access / API	Publishing Program	Journal	*Library Publishing Program*
Archiving	Perpetual access	Library	Academic institution	*Repository Program*

Figure 11.2. Functional affinity between repository- and library-based publishing programs.

publishing programs developed and administered within academic libraries—the same setting where repository programs frequently are established. If we apply the analysis criteria identified therein to library-based publishing programs, the convergence with the suggested repository-based disaggregated model becomes more apparent (see Figure 11.2).

Certain processes, such as *perpetual access,* completely converge, being components of the inherent mission of academic libraries. Other processes are apparently distinct. However, when abstracted from their contingent implementation they manifest shared essential characteristics. While the notion of *overlay journal* has yet to emerge as a competitive alternative to the established publishing system, a similar approach can be adopted locally at individual institutions by aligning and integrating library-based publishing and repository programs.

Processes in Repository and Publishing Programs Within Libraries

Repository programs and publishing programs within academic institutions can be understood in terms of processes that, combined, describe typical workflows within them.

Repository Programs

Consider the repository perspective first. We can describe the workflow of an institutional repository infrastructure according to the following macro-level managed processes. No assumption is made on how these processes are implemented at the technology level.

Submission: The process by which new content is submitted to and received by a repository.

Review: The process by which submitted content is assessed against eligibility criteria and accepted in the repository. Criteria include but are not limited to fitness to the collection policy and intent as well as quality assurance on the submitted item and the associated description. Such a process may be enabled entirely by policy (e.g., any item submitted by an eligible community member may pass *review*).

Distribution: The process by which content accepted into a repository is made available online to the intended audience.

Curation: The ongoing process of ensuring the persistent access and availability of content admitted into a repository, including but not limited to, routine audit, metadata remediation, infrastructure maintenance, and format migration.

Aspirationally (if not always functionally), repositories fulfill both a *Publication* and a *Stewardship* role within academic institutions (i.e., they are meant to provide persistent access to their content for the future). Therefore we included here a *Curation* process.

Library Publishing Programs

The workflow of individual journals within a library publishing program can be effectively described, appealing to similar, if not identical macro-level managed processes:

Submission: The process by which new content is submitted to and received by a journal.

Review: The process by which submitted content is assessed in scope, quality, and form, ending with a publishing decision over submitted content. Review is an iterative process that may account for a number of editor-, review-, and author-introduced revisions. Peer review is a component of *Review*. Production (the process of preparing content for publication, including but not limited to copyediting, formatting, typesetting, etc.) is as well.

Distribution: The process of publishing the content in a form intended as the final authoritative one for the journal.

Curation: The ongoing process of ensuring the persistent access and availability of the published content, including but not limited to routine audit, metadata remediation, infrastructure maintenance, and format migration.

The similarities with the repository processes presented above, in particular when considering mediated repositories, is not only in the common terminology adopted here; the essence of the described processes is very much the same if we abstract from the contingencies of how these processes are instantiated and the potentially different actors involved. Library publishing programs provide some level of *stewardship* over their content (part of which is involved in the iterative *Review* process), but not necessarily to the level expected by the mature digital stewardship program where content is curated for the long term. Nevertheless we include here the *Curation* process as well, with the expectation that mature library publishing programs would act to ensure the digital longevity of their published content.

The specifications for the high-level processes inherent in repository systems and journal publishing programs are similar enough to become indistinguishable at the program level. In both workflows, content to be published follows a process of submission, review, and preparation prior to publication. The functions inherent to each process step, the sufficiency criteria applied, and the agents conducting the assessment and performing the functions necessarily differ. For example, a fairly traditional journal publishing *Review* process involves a series of communications between editors and reviewers in the discussion of specific criteria applied to the

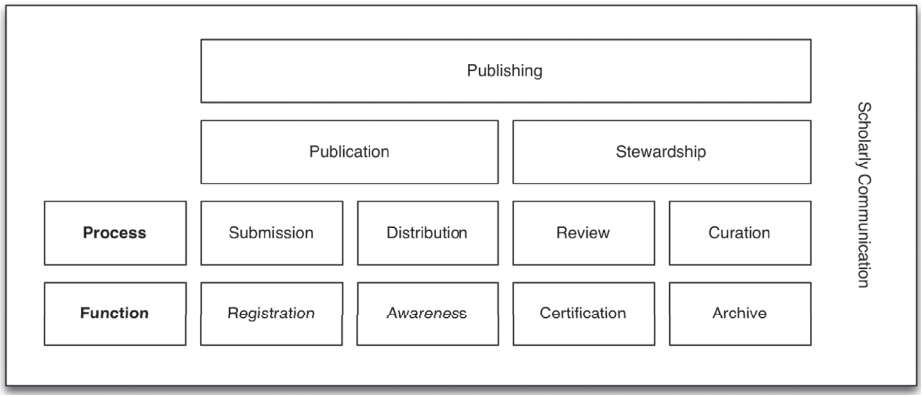

Figure 11.3. Assessment framework modeling functions and processes.

submission before arriving at a publication decision. In the repository, such an "academic quality" review might be covered by: (a) submission criteria that admit only postprints of accepted journal articles; (b) a collection policy that permits administrators to act as editors and curators over subsections of collected content; or (c) a single repository administrator acting upon ad hoc publication criteria.

AN EMERGING ASSESSMENT FRAMEWORK

The analysis presented so far allows us to derive an assessment framework defining the requirements to successfully fulfill the core functions of scholarly communication in terms of processes (see Figure 11.3). This framework applies an integrated perspective that considers both a repository program and a journal publishing program within a library.

The approach taken here models the activities that are required to enable the core functions of scholarly communication in terms of the aforementioned processes. The registration and awareness functions together describe the minimum requirements for communicating research and scholarship and correspond to *Publication* in the model. The core processes that instantiate *Publication* are *Submission* and *Distribution*. The *Publication* activity is, however, understood as a component of a broader *Publishing* enterprise, which also includes *Stewardship* of submitted content. *Stewardship* is instantiated by the processes of *Review* and *Curation* where content is iteratively assessed, refined, accepted for publication, but also

where recurring auditing activities ensure its perpetual access and provide a long-term perspective on the issue of digital longevity, closing the circle of scholarly communication.

When considering a repository and publishing program, this framework can be applied at multiple levels for assessing the following:

1. The capabilities of an organization to satisfy the basic requirements for effective scholarly communication
2. The contextual feasibility and benefits of integrating a publishing program with a repository program
3. The modularity and extensibility of technical infrastructures

Applying this framework to existing or envisioned scenarios allows stakeholders to programmatically assess their programs in place. The previously presented case study at Columbia is provided as an example.

An Application Example

The Columbia University case study presented earlier in this chapter provides an example scenario where this framework can be applied to assess the integration feasibility of a mediated repository program such as Academic Commons and a mature publishing program such as the one carried on at CDRS.

Publication of content on both Academic Commons and in many of the journals that are managed within the CDRS Publishing Program involves the following managed processes: *Submission, Review,* and *Distribution*. Despite being implemented differently—in terms of both adopted technology and practical procedures—this convergence provides the common ground for evolving our *publishing* enterprise into a more integrated infrastructure.

When considered from this analytical perspective, the integration process can be decomposed into components that reflect the identified processes, each individual one addressed (potentially) at different stages. Academic Commons is also intended to provide long-term digital preservation capabilities, *de facto* implementing the *Curation* process. Right now some journals within the publishing program submit their content to Academic Commons, but this light integration leaves open a series of issues, including

the replication of content between platforms, with the inevitable issue of version and variance management. A tighter integration would completely leverage the Fedora component of Academic Commons not only as a preservation infrastructure, but also as the infrastructure to provide access (though the individual journal front-ends) to the authoritative copy of each article.

The modular and layered infrastructure of Academic Commons is already suited to support not only multiple distribution channels, but also multiple submission channels, supporting a tighter, yet flexible integration of both the *Distribution* and the *Submission* processes via module extensions. The focus of this high-level assessment is to show how the framework can be leveraged to model real situations and break down an infrastructure into more manageable process-based components. The situation at CDRS can be described as a scenario with a *nonmediated repository* and a *publishing program*. Depending on the specific setting and contexts of an institution, other prototypical scenarios can be identified and analyzed according to the framework.

TOWARD AN INTEGRATED MODEL OF INSTITUTIONAL PUBLISHING

The proposed integrated model of institutional publishing suggests opportunities for libraries looking to advance both publication programs and persistent access and preservation repositories. For many existing programs, this is evident: either the selected repository platform promotes these possibilities out of the box (e.g., Digital Commons and DSpace), or the practical constraints around resourcing scholarly communication programs require the flexibility to apply staff and infrastructure to multiple service approaches.

But even for those institutions where journal publishing and digital repository programs have matured largely independently, such as at Columbia, the rationale for adjusting the program development roadmap toward purposeful integration becomes apparent. In the integrated view, the repository becomes the publishing platform, both in the outreach and communication and in the approach to platform development. Language matters, however, in outreach to authors and depositors about the availability of new library programs and services. Despite the alignment of function and process, program managers may prefer differentiation between the useful concepts “deposit”

and "publish" to direct contributing authors to multiple service entry points, and still the benefits of observing the integrated model may persist.

Program integration may serve to enhance the scalability of services and to maximize the efforts of limited staff working with the publishing platform. This may be at odds at times with the spirit of experimentation and flexibility around the business model and customized application development approach to journal publishing in libraries. Does the integrated model of program development therefore pose a threat to the core value propositions and differentiating factors for journal editors participating in such programs?

Underlying all of this speculation, of course, rests the presumption that institutions will choose to continue resourcing a shift in scholarly publishing infrastructure in ways that bring capacity and expertise in-house, returning control of a once arcane and print-based process to authors and the universities that support their work. Integrated publishing programs within libraries lay the necessary groundwork for viable, complementary alternatives to traditional publishing and archiving scenarios. Covering most of the components of the scholarly communication workflow, repository programs have demonstrated that commitment to the requisite infrastructure; of particular note are the extensible platforms that have resulted from sustained, coordinated multi-institutional, volunteer, and consortial efforts. Through publishing programs in libraries, the remaining essential components come into view, aided again by formal, cross-institutional initiatives that foster publishing production expertise among library staff. The barriers to introducing manageable, cost-efficient options for publishing scholars through the proliferation of library-led repositories at the programmatic level are few and dwindling.

NOTES

1. See //http://www.dspace.org.
2. See http://openjournalsystems.com.
3. See review by Newton et al. https://authorea.com/users/6729/articles/7032/_show_article#article-paragraph-Literature__space__Review__dot__md
4. See http://acrl.ala.org/newroles/?page_id=263 for Barbara Fister's excellent overview as contributed to ACRL's New Roles for the Road Ahead (2015) on advancements in this area.

5. See http://www.openarchives.org.
6. The idea of "overlay journals" has been recently revamped within the humanities community. See https://www.openlibhums.org/2014/04/07/olh-overlay-journals/.
7. Center for Digital Research and Scholarship; see http://cdrs.columbia.edu.
8. See http://library.columbia.edu.
9. See http://academiccommons.columbia.edu/.
10. See http://fedorarepository.org/.
11. This summary is adapted from Van de Sompel and colleagues' "Rethinking Scholarly Communication," http://www.dlib.org/dlib/september04/vandesompel/09vandesompel.html
12. See http://www.lockss.org/.
13. See http://www.clockss.org/.
14. See http://www.portico.org/.
15. See http://www.iso.org/iso/catalogue_detail.htm?csnumber=56510.

REFERENCES

Bonn, M. (2014, March). Tooling up: Scholarly communication education and training. *College & Research Libraries News, 75*(3), 132–135.

Crow, R. (2002). *The case for institutional repositories: A SPARC position paper*.

Lippincott, S. K. (Ed.). (2014). *Library publishing directory 2015*. Library Publishing Coalition.

Lynch, C. A. (2003). Institutional repositories: Essential infrastructure for scholarship in the digital age. *portal: Libraries and the Academy*, *3*(2), 327–336.

Newton, M. P., Cunningham, E. T., & Morris, J. N. (2013). Emerging opportunities in library services: Planning for the future of scholarly publishing. In *Library publishing toolkit* (pp. 109–117). IDS Project Press.

Perry, A. M., Borchert, C. A., Deliyannides, T. S., Kosavic, A., & Kennison, R. (2011, September). Libraries as journal publishers. *Serials Review, 37*(3), 196–204.

Roosendaal, H. E., & Geurts, P. A. T. M. (1997). *Forces and functions in scientific communication: An analysis of their interplay*. Conference on Co-operative Research in Information Systems in Physics, September 1–3, 1997, University of Oldenburg, Germany (pp. 1–32). Retrieved from http://doc.utwente.nl/60395/