

*Prepublication copy submitted to Facet Publishing 16 September 2013*

## 8 The Prospects of Open Access Repositories

**Karen Calhoun**

Cornell University Library (retired)

[ksc10@cornell.edu](mailto:ksc10@cornell.edu)

**Note:** This is a preprint of a chapter whose final and definitive form was co-published in *Exploring Digital Libraries: Foundations, Practice, Prospects* by [Facet Publishing](#) (2014) and [ALA Neal-Schuman](#) (2014).

### Overview

This chapter focuses on the potential of open access repositories for having a distinctive positive impact on scholarship and, more broadly, on their prospects for increasing the social and economic value of digital libraries. In addition to extending chapter 4's discussion of open access repositories into new territory, it relates the frameworks presented in chapters 6 and 7 to this particular type of digital library. Topics include subject-based and institutional repositories and their value; issues around recruiting repository content, including deposit mandates; legal frameworks, copyright and open access; discipline-specific norms, practices and reward systems; the discoverability of scholarly content; sustainability of repositories; e-research data management; and prospects for the emergence of a global ecosystem of repositories.

### Successful subject-based repositories

The most successful subject-based repositories have grown organically around the scholarly communities they serve (see the examples in chapters 2, 4, 6 and 7), and they are woven into the way their disciplines communicate. As Erway (2012) notes in her review of several thriving subject repositories, “the central repository for a researcher’s field of study is where he goes for information, to see what’s been published, and to look for collaborators. It’s only natural that he would think of the same location when it comes time for him to deposit his work.” Due to their

**Keywords:** Institutional repositories—Evaluation; Success factors; Scholarly communication; Open access; Copyright; Self-archiving; Deposit mandates; User-centered design; Google and Google Scholar; Discoverability; Web search engines; Research—Information services

firm foundations within communities of research and practice, the successful subject repositories have tended to attract more submissions than institutional repositories.

Adamick and Resnik-Zeller wrote two articles about subject repositories (2010a; 2010b). The second article profiles ten selected subject repositories across nine metrics: year founded, subjects covered, software, content types, deposit policies, copyright policies, hosting, funding and governance. Overlaying Iriberry and Leroy's 2009 life-cycle model of online communities (figure 7.1) on the profiles of these successful subject repositories, it is clear that the repositories have evolved through the life cycle, from a strong community orientation at inception to a high degree of trust and participation at maturity. As a result, while Google is an important referrer of searches to subject repositories, the most successful repositories can be said to be destination sites (sites that people visit by going directly to their URLs through typing or bookmarks). As discussed later, this has implications for repository interface design and development.

### **The value of institutional repositories**

Several of the digital library experts interviewed for this book noted that some institutional repositories have been built without an understanding of research, teaching and learning practices. As a result, the articulated benefits of these repositories align and resonate poorly with the needs of contributors and the hosting institution. Along these lines Sefton and Dickinson (2011) note "there is more work to be done to align the library-centric view of institutional repositories with their uses in other academic contexts." These "contexts" comprise the total package of value delivered to parent institutions, faculty and repository end users.

The following analysis of the digital library literature supports interviewees' reservations about institutional repositories. Among the major barriers to the success of institutional repositories are:

- A lack of clarity around purpose and focus
- Weak understandings of community needs and attitudes
- Scholars' lack of awareness of the repository or its benefits
- Recruiting content

Iriberry and Leroy's analysis (see chapter 7) indicates that if a network-based service's intended communities do not actively engage and participate, the service will die. The aims of this section are to offer ideas for making institutional repositories more successful by: (1) enumerating some priorities for aligning repositories around the communities to be served; and (2) better articulating the value of repositories.

### ***Purpose, focus and community needs***

Getting attention for an institutional repository is challenging, and understanding what the repository's stakeholders and target audiences want and need is even more difficult. Oya Rieger (2007), a digital library expert who provides oversight for a number of repositories, recommends gathering stakeholders and conducting a repository needs assessment to understand the organizational environment and existing workflows, identify common ground and use cases, and generate stakeholder awareness. Rieger is writing of the process of selecting a repository for an institution, but her advice around conducting a needs assessment is equally applicable when the repository is already in place and course corrections are needed. Unfortunately, in a census of institutional repositories, Markey and others (2007, 31) found that systematic needs

assessment has not been a significant factor in libraries' decisions to start or maintain institutional repositories.

### ***Awareness and recognition***

Raising awareness, recognition and branding of institutional repositories are priorities. The low level of faculty awareness of repositories has been reported for some time (for example Swan and Brown 2004, 220; Rowlands and Nicholas 2005; Watson 2007; Morris and Thorn 2009), so Gale Moore's 2011 report of low faculty awareness at the University of Toronto is not a surprise. In the Moore survey, less than half (43%) of faculty respondents were aware of the university's institutional repository, which had been in place for seven years at the time of the survey. Of those who had heard of it, only 15% had deposited work there. Further, over three-fourths of respondents were unaware of the university's open journal and book publishing services, and two-thirds were unaware of services related to hosting media and archiving data sets.

### ***The problem of recruiting content***

Institutional repository managers can be hard pressed to articulate the value to faculty and researchers persuasively enough to motivate the contribution of content. Consider one interviewee's remarks following a set of conversations recently conducted with faculty members:

Faculty members think a repository and open access are good ideas and the right things to do. However, asked if they would deposit their work in the repository, they said no, they would not want to do this work.

### ***Faculty perceptions***

The digital library literature suggests that faculty and researchers are not only unwilling to do the work of depositing content in institutional repositories—many do not see sufficient reason to do so. One of the digital library experts interviewed for this chapter, who had recently been talking

with faculty about institutional repositories, noted faculty members' prevalent belief (probably false) that the articles they write are already being found by all of those who are interested in them. Troll Covey's research (2011, 8) also indicates that some faculty may not believe that visibility and access to their work are problems, and thus there is little need to invest time in self-archiving. Looking at faculty attitudes from the perspective of scholars as readers (instead of as writers), Swan and Brown's results suggest nearly two-thirds of faculty believe they have easy access to most or all of the articles they need for their work (2005, 13). They found some variability in these results, with humanities scholars reporting somewhat more difficulty accessing what they need. Even when faculty and researchers do make open access copies of their published content available, a study by Kim (2010, 1914) found a preference for linking to open access copies from their personal web pages (66.7% of respondents), followed by research group and departmental websites, then subject and institutional repositories.

#### *Value propositions*

Another interviewee remarked "libraries have not put forward adequate value propositions for the repositories they host." Numerous studies, starting with one at the University of Rochester River Campus Libraries (Foster and Gibbons 2005), have linked the problems with recruiting faculty content to the way that librarians talk with faculty about repositories. Issues with content recruitment have continued: six years after the Foster and Gibbons study, Wacha and Wisner (2011) studied 45 institutional repositories and found that only three contained the highest impact articles of faculty at their institutions. Chapter 4 presents the extensive evidence for low deposit rates in institutional repositories; the literature review of Lercher (2008) is also a good source of evidence for low deposit rates. The issue is what institutional repository builders and managers can do about it.

What follows are some solutions based on better articulating the value of institutional repositories for faculty and researchers. Conducting needs assessments—and then acting on what is learned—has already been mentioned. Formulating a crisp value proposition for faculty is also important. Two reports, noteworthy for their community-centered approaches, are particularly sources of ideas. The report from the University of Toronto has already been mentioned (Moore 2011, 130-131). The other is a Mellon-funded pilot study from Palmer, Tefteau and Newton (2008), which featured interviews with faculty, library liaisons, and repository managers. While Moore's report looks into faculty attitudes and practices, Palmer's report focuses on what problems that institutional repositories might solve for faculty. Table 8.1 synthesizes the findings of the two reports into a potential value proposition for faculty.

Another source of ideas for securing commitment, recruiting content, and generally improving institutional repositories for a variety of audiences is Bell and Sarr (2010) who report how the University of Rochester's institutional repository was re-engineered. They added support for research and writing workflows, collaboration with peers, usage statistics and author profiles. In addition, the many reports and articles produced by the IMLS-funded MIRACLE project (Making Institutional Repositories A Collaborative Learning Environment; [miracle.si.umich.edu](http://miracle.si.umich.edu)) can guide repository managers' planning.

### ***Articulating the value***

Table 8.1 collects content from a number of sources that have reported evidence-based results suggesting the value of institutional repositories to different stakeholders and target audiences. Subsequent subsections discuss these sources. In particular, Alma Swan (2008) provides an in-depth, detailed analysis that articulates the value of digital repositories. She also defines a typology of business cases for them. In more recent work, Swan (2011) identifies the benefits of

repositories to various audiences, including their value supporting national research assessments by providing a record of institutional research outputs.

**Table 8.1 The value of institutional repositories to stakeholders and target audiences**

Stakeholder or target audience	Value for this stakeholder or audience
<b>Stakeholder: Hosting library</b>	<ul style="list-style-type: none"> <li>• Removing barriers and fostering open access to scholarship</li> <li>• Raising the profile of the library’s curatorial and facilitation roles in scholarly communication processes</li> <li>• Raising the profile of the library’s role in showcasing research at its parent institution</li> <li>• Demonstrating how the library contributes to advancing the institutional mission and goals</li> </ul>
<b>Stakeholder: Parent institution</b>	<ul style="list-style-type: none"> <li>• Showcasing the institution’s intellectual output and raising prestige</li> <li>• Providing a source of metrics for institution-level scholarly outputs</li> <li>• Helping the institution to demonstrate its value to its communities and funders</li> <li>• Providing the means to publish and provide discovery and access for other types of intellectual and cultural assets produced at the institution (e.g., teaching materials, student honors work, working papers, presentations, conference proceedings, etc.)</li> </ul>
<b>Audience: The institution’s end users</b>	<ul style="list-style-type: none"> <li>• Finding out what research is being conducted locally</li> <li>• Collecting information related to institution-based dissertation topics and honors theses</li> <li>• Networking – finding people in different departments or potential collaborators</li> <li>• Getting to unpublished content not available elsewhere (including data, video clips, learning materials, content related to events, etc.)</li> <li>• Finding institutionally relevant primary source documents for use in teaching and writing</li> <li>• Promoting research done within the institution to institutional colleagues</li> </ul>

Stakeholder or target audience	Value for this stakeholder or audience
<b>Audience: the institution's faculty and researchers</b>	<ul style="list-style-type: none"> <li>• Increasing exposure of an individual scholar's forthcoming and already published work (pre-prints and post-prints)</li> <li>• Providing exposure for a scholar's unpublished work (like working papers)</li> <li>• Solving specific information visibility, management, or access problems (these vary by discipline)</li> <li>• Supporting specific faculty workflows for managing and disseminating digital content</li> <li>• Coordinating work with existing and emerging subject-based or funder repositories</li> <li>• Offering publishing services for a spectrum of the institution's intellectual output</li> <li>• Attracting audiences for content that is not easily discovered in the corpus of scholarship</li> <li>• Broadening dissemination of academic research to the public</li> <li>• Contributing to the open access movement</li> </ul>
<b>Stakeholder: Government agencies</b>	<ul style="list-style-type: none"> <li>• Supporting national research assessments</li> <li>• Demonstrating the societal benefits of publicly funded research</li> <li>• Supporting knowledge transfer and economic growth</li> </ul>

### *Hosting libraries*

Many institutional repository managers can articulate the value of repositories to their own libraries. Among other writers, Bankier and Smith (2010) and Markey and others (2009) observe the benefits to hosting libraries; table 8.1 lists these.

### *Addressing the serials crisis*

An additional hoped-for benefit of institutional repositories for libraries has been to lower the costs of access to highly-priced scholarly journals (in other words, to address what is called the "serials crisis"). An early hope was that a robust system of open repositories would address spiraling journal costs in addition to delivering other benefits. Over the past ten years, that hope has been tested, and it remains unclear whether open access models (that is, open access journals and repositories) will reduce the costs of access to the scholarly literature (see



Waijers 2008). Burns and others' consideration of the evidence suggests "it is much too early to tell what kind of overall financial savings, if any, these entities [institutional repositories] have created" (2013).

#### *Parent institutions*

Table 8.1's articulation of the benefits of an institutional repository to a library's parent institutions is drawn from Bankier and Smith's 2010 analysis.

#### *The institution's end users*

As part of their work with the MIRACLE project, St. Jean and others (2011) examined the value of institutional repositories to end users. There are at least three ways to approach a consideration of the value of institutional repositories to end users. Chapter 4 of this book covers the high visibility and use of repository content by end users around the world as a result of indexing by Google and Google Scholar. Chapter 6 covers the potential value of repositories to end users as members of society. In contrast to these two perspectives on end users, St. Jean and her colleagues studied end users who were deliberately using the institutional repository as a local resource and destination site for locally-based end users. As they point out, not much research has been conducted on this particular end-user perspective. Their results suggest another avenue for aligning an institutionally-based repository to benefit an institutionally-based community. Table 8.1 lists the value of the institutional repository for this category of end users, based on their end-user interviews.

The *personas* research of Maness, Miaskiewicz and Sumner (2008) at the University of Colorado at Boulder should also be mentioned in the context of the institutional repository's value for end users. Personas are concrete but fictitious representations of a group of target users with common characteristics. A principal finding contradicted prior assumptions about

target audiences' goals: the authors had expected a desire for a place to deposit pre- or post-prints of published work. Instead, the personas wanted the repository to be a place to share teaching and learning materials, identify potential collaborators, and promote their research to institutional colleagues. Maness and his colleagues' findings reinforce the results of St. Jean and suggest a value proposition specifically targeted to the institutionally-based, local community.

### *The institution's faculty*

As already noted, the list in table 8.1 is drawn from the author's analysis of studies by Moore (2011) and Palmer, Teffeau and Newton (2008), in addition to recent work by Swan (2011). With respect to the qualification of stressing open access in the value proposition to some but not all faculty and researchers, Palmer's study (26) found that while many faculty support open access, some have concerns (e.g., loss of control, timing, versioning, quality). In general the study found that gaining faculty engagement was more about explaining how the repository solves particular problems that faculty or researchers encounter in their daily work (22).

### *Government agencies*

Increasingly, governments and international funding bodies are taking a keen interest in tracking and understanding the real-world societal benefits being produced by publicly funded research. Examples include the European Commission, the UK's Funding Councils and Research Councils UK; a following section discusses these.

## **The policy and legal frameworks**

### ***Open access and self-archiving***

#### *The open access movement*

The open access movement has its roots in the first decade of digital library research and practice (see chapter 2). As noted in chapter 2, participants in the open access movement advocate making scholarly information publicly available in a way that is “digital, online, free of charge, and free of most copyright and licensing restrictions” (Suber 2004). Formal open access declarations that were made at meetings in Budapest ([budapestopenaccessinitiative.org](http://budapestopenaccessinitiative.org)), Bethesda (Suber 2003) and Berlin (Max Planck Society 2003) form the backbone of the movement’s principles and objectives. Denicola (2006, 353-354) provides a brief summary of the beginnings of the open access movement.

#### *Recent developments in Europe*

As discussed later in this chapter’s section on deposit mandates, the US National Institutes of Health (NIH), and many universities and funding agencies worldwide have issued policy statements requiring that papers produced with public funding be made openly available to the public. More generally, the UK’s Finch Report (Finch and Jubb 2012) and the endorsement of its recommendations by the UK government, UK Funding Councils and Research Councils UK, and Research Excellence Framework will have the effect of placing great emphasis on open access to UK scholarly publications, especially after 2014 (Research Information Network 2013). The European Commission’s communication of new policies for open access to publications and research data (2012), with its target for establishing open access policies in all member states by 2014, has further magnified the fresh momentum for open access to research that has been created around the world by the Finch Report.

### *Green and Gold open access*

With respect to the processes of scholarly communication, the open access movement has evolved in two directions: “Gold” and “Green” open access. The literature that discusses open access is large, spirited and sometimes divisive, so much so that in October 2012, IFLA and the International Publishers Association issued a joint statement calling for a more nuanced, empirically-based debate ([ifla.org/publications/enhancing-the-debate-on-open-access](http://ifla.org/publications/enhancing-the-debate-on-open-access)).

Two articles, one by Stevan Harnad and others (2004) and one by Jean-Claude Gu  don (2008) provide frequently-cited perspectives on Green and Gold open access. A more recent look at this ongoing debate is from Shieber (2009) and Gargouri and others (2012). A report from the European Commission discussing new open access policies (2012, 5, 7) provides brief definitions of the Green and Gold methods.

In brief, Gold open access is associated with publishing open access journals. Green open access is associated with self-archiving, which involves authors’ depositing copies of their content on their own or group websites or in open access repositories, often after an embargo period (a delay before full text is openly available to all audiences). Self-archiving sits alongside the traditions of scholarly publishing. It is not a replacement for what publishers do. Because of its ties to the subject of this chapter—open access repositories—Green open access and self-archiving are the focus in this book (not Gold).

### ***Copyright and repositories***

#### *Scholars’ copyright concerns*

In 2002, 90% of scholarly journal publishers required that authors transfer copyright to the publisher, according to a survey of eighty UK and US journal publishers (Gadd, Oppenheim and Proberts 2003). Since then the percentage of publishers requiring copyright transfer has been

declining (Cox and Cox 2008), and an increasing number of publishers now allow some form of self-archiving by authors. Nevertheless many scholars continue to transfer the copyright in their content to publishers (or to believe that they have). It is not surprising then that scholars who are considering depositing their work in an open access repository are concerned about violating publisher copyright agreements (Swan and Brown 2005; Watson 2007).

Other studies suggest that scholars are concerned, confused or indifferent to copyright issues. Kim's (2010) and Troll Covey's (2011) suggest that copyright concerns are significant barriers to faculty willingness to self-archive. Morris' (2009) and Moore's (2011) surveys found that many faculty members do not understand publisher contracts and copyright terms, and few have a clear picture of what rights they retain or could retain. In addition the Morris study found that authors overestimate their self-archiving rights for published PDFs and underestimate what publishers allow them to do with pre-prints and post-prints. Finally, it is possible that some scholars simply do not care much about copyright. Nine years ago, Rowlands, Niccolas and Huntington (2004, 265) found that 46% of surveyed authors "took no interest at all" in copyright matters. These and other attitudes translate to major deterrents to self-archiving and thus to depositing content in open access repositories.

#### *Copyright, authors' rights and self-archiving*

Copyright law varies from country to country and is extremely complex. In brief, and acknowledging the possibility of variations in different university settings, academics who write a scholarly articles are generally viewed as copyright holders, and they can frequently choose the terms under which articles are distributed or reused (although scholarly tradition plays a role in what choices are truly actionable). Traditionally, in university settings, the author of a scholarly article has transferred his or her copyright to a publisher in exchange for what the publisher adds to the process: managing the peer review process, producing and disseminating the

journals in which articles are published, and so on. Academic authors of articles in university settings still frequently sign copyright agreements that transfer their copyright to the publisher, but as of this writing it is not unusual for scholars to negotiate these agreements to either retain some rights or retain copyright but license some rights to the publisher. One possibility is for the author to grant the publisher the right to publish the article, but keep the right to make the article openly available under specific conditions, for example in an open access repository.

Sometimes these agreements are subject to embargo periods.

In a survey of scholarly authors conducted in 2004, Swan and Brown (2005, 56-57) asked who held the copyright to the last article the author self-archived. Their results were that 35% of authors claimed to hold copyright themselves, 37% assigned copyright to the publisher, 22% didn't know who held the copyright, and 6% assigned copyright to some other party (e.g., employers). Asked if they were required to ask the publisher's permission to self-archive, 47% said no, 17% said yes, and 36% did not know. Asked if they did ask the publisher's permission to self-archive, 84% said no. In a 2008 study of thousands of self-archived articles on faculty web pages, Troll Covey (2009, 240) found that "38 percent are not aligned with [publisher] policy in terms of whether self-archiving on personal and departmental Web sites is allowed and whether publisher policy allows, prohibits, or requires self-archiving the publisher PDF. "

#### *Sherpa RoMEO and copyright clearance*

A number of organizations that advocate open access provide information for scholars who want to retain self-archiving and other rights (see for example the SPARC Author Rights Addendum, [www.arl.org/sparc/author/addendum.shtm](http://www.arl.org/sparc/author/addendum.shtm), and the Science Commons Scholar's Copyright Addendum, [scholars.sciencecommons.org](http://scholars.sciencecommons.org)). Sherpa RoMEO ([sherpa.ac.uk/romeo](http://sherpa.ac.uk/romeo)), based at the University of Nottingham in the UK, is a service that maintains a searchable directory of publishers' copyright conditions for self-archiving. The services tracks "green"

publishers and journals and other information needed by authors who wish to self-archive. Sherpa RoMEO has been a highly beneficial service to open access repository managers who help authors to deposit their work. Hanlon and Ramirez (2011) describe how repository managers use Sherpa RoMEO and other tools to support their copyright clearance activities. Unfortunately, faculty awareness of Sherpa RoMEO may be low, as reported in the results of the Troll Covey (2011) focus groups: not one of the hundred focus group participants had heard of Sherpa RoMEO.

## **Deposit mandates**

### ***Current situation***

Policies requiring that researchers make their papers available via open access in a particular repository or repositories are known as “deposit mandates.” Specific deposit mandate policy terms can vary, and there are a number of kinds of deposit mandates, with the main ones being those of funding agencies, institutions or groups of institutions, and university departments or other sub-units. An important practical influence on the current situation came in 2008, first when the US National Institutes of Health (NIH) announced a new public policy requiring that papers from NIH-funded projects be submitted to PubMed Central and made publicly available within 12 months of publication ([publicaccess.nih.gov](http://publicaccess.nih.gov)). That same year, the European Research Council announced it would require papers and monographs supported by Research Council funding to be made publicly available no later than six months after publication (for the latest version of the policy see [erc.europa.eu/documents/open-access-policy-researchers-funded-erc](http://erc.europa.eu/documents/open-access-policy-researchers-funded-erc)). NIH and the European Research Council were not the first funding agencies to adopt such policies; for example the deposit mandate of the Wellcome Trust in the UK had a major impact when it was implemented in October 2005 (see the timelines in the Open Access Directory, [oad.simmons.edu/oadwiki/Timeline](http://oad.simmons.edu/oadwiki/Timeline)). A number of other funding agencies and scholarly institutions followed suit. At the time of this writing, over 250 organizations worldwide have

registered deposit mandates in ROARMAP ([roarmap.eprints.org](http://roarmap.eprints.org)), which maintains a searchable list of these.

Peter Suber, a leading open access advocate in the US, quickly placed the NIH and European Research Council's announcements in the context of other open access deposit mandates around the world (2009). He also wrote of the mandates adopted in 2008 by Harvard University's Faculty of Arts and Sciences and the Harvard Law School. Both attracted a good deal of attention when faculty members voted unanimously to implement open access deposit mandates. Harvard was not the first university in the world to adopt mandates; 11 other universities already had deposit mandates at department or university-wide levels, with Queensland University of Technology taking the lead with its university-level mandate in September 2003 (see [oad.simmons.edu/oadwiki/Timeline](http://oad.simmons.edu/oadwiki/Timeline)).

### ***History of deposit mandates***

Open access evangelist Stevan Harnad has vigorously argued the merits of open access deposit mandates. Stephen Pinfield (2005), another early advocate for deposit mandates, supports them on the grounds that they will accelerate the widespread adoption of open access, thereby improving the process of scholarly communication, increasing the impact of individual papers, and maximizing the free availability of large numbers of high quality scholarly content. In his study of seven Australian universities with institutional repositories, Arthur Sale (2006) found that mandatory deposit policies result in more self-archiving than voluntary deposit policies. In a later article, Sale (2007) recommended that repository managers pursue a "patchwork" (department-level) mandate as a quicker route to growing institutional repository deposits.

Swan and Brown's report on author self-archiving is frequently cited for its finding that over four-fifths of authors would willingly comply with a deposit mandate (2005, 62-63). Carr and others



(2006) also urged institutions and funders to seriously consider mandating the practice of self-archiving. Thomas and McDonald's comparative analysis of voluntary- and mandatory-deposit repositories (2007) led them to suggest that deposit mandates might increase the number of papers deposited per author. More recently, a comparative study of two Australian universities by Mary Anne Kennan (2011) found that one of the universities' open access and repository advocacy, education and support, combined with deposit mandates, did result in more repository deposits and deeper faculty engagement with the issues around open access publishing.

### ***Mixed reviews of deposit mandates***

There are other perspectives on deposit mandates, and whether a deposit mandate exists or not, self-archiving has been difficult to promote and manage. Compliance with the NIH and Wellcome Trust deposit mandates has developed slowly. Poynder (2012) reported that six years after putting its deposit mandate in effect, the Trust was achieving a 55% compliance rate, prompting the Trust to take new steps to enforce the mandate. Poynder's investigation suggested that the NIH mandate was getting better results, with a compliance rate of 75% after four years.

In an article formatted as a debate on the criticality of mandates to institutional repository success, with Harnad taking the affirmative and Nancy McGovern the negative (2009), McGovern pointed out issues with the expense and difficulty of promulgating and enforcing mandates. Arguing that incentives other than mandates (such as value-added services), peer pressure, and evidence of positive outcomes might contribute more to institutional repository success, McGovern concluded "mandates alone – or possibly at all – are not the determining factor in the success of institutional repositories." The aggregated findings of a number of other studies support McGovern's conclusion:

- Kim's study (2010, 1916) of what motivates faculty to deposit their work in repositories suggests that the most powerful motivator is altruism: support for the spirit of open access to help others (that is, by implication, not mandates).
- Moore's results (2011, 4) confirm that faculty are generally in favor of the *principles* of open access, but her study also suggests that faculty understanding of open access *in practice* is limited, and strongly shaped by the norms of each discipline (again, by implication, not mandates).
- Results of a faculty survey and focus groups conducted by Creaser and others (2010, 156) indicated mixed reactions to mandates and led to questions about their effectiveness for motivating faculty self-archiving.
- Palmer and others found faculty have mixed views about mandatory deposit in institutional repositories, with one faculty member noting "there are lots of things that are mandated and don't happen on campus" (2008, 26-28).

Other incentives may help to drive self-archiving other than, or in addition to mandates. Ferreira and others (2008) discussed an original—and highly successful—approach that combined financial incentives for departments ("carrots") with the "stick" of an institutional deposit mandate. Other authors point to incentives in the form of value-added services, better alignment with researcher workflows, help with copyright clearance, and more, as discussed in other sections of this chapter. Bankier and Smith (2010), while noting that some deposit mandates have been successful, also point out that such mandates continue to focus institutional repositories narrowly on pre- and post-prints, thereby ignoring opportunities for the repository to serve other purposes for other audiences. Table 8.1 lists these audiences and opportunities.

## **Other issues with self-archiving**

### ***Scholarly tradition and discipline-specific reward systems***

In addition to copyright and other concerns, Kim's 2010 faculty survey also identified the additional time and effort to deposit content as deterrents to faculty self-archiving. Others might add the traditions of academic reward systems to this list of barriers. In fact, faculty complaints about the additional time and effort required to self-archive in an institutional repository may to some degree be a presenting symptom of a deeper issue: faculty feel comfortable with their own disciplines' arrangements for disseminating new intellectual content and lack sufficient motivation to change.

In particular, when faculty members perceive a lack of discipline-based rewards (or worse, the possibility of risk) associated with depositing intellectual assets in the institutional repository, attempts to modify their behaviors and choices will be uphill battles. A variety of discipline-specific traditions and value systems exist, making this particular barrier even more complex to overcome, because the battle must be waged on multiple, discipline-specific fronts. While the processes of scholarly communications are changing, not all disciplines are changing at the same rate, and it is a mistake to underestimate the weight of discipline-specific norms, attitudes, fears or concerns. Moore's study found that faculty are willing to explore alternatives, but their choices of where and how to publish their work continue to be driven primarily by reputational factors such as peer review and journal readership/impact (2011,4).

Oya Rieger's social constructivist analysis of institutional repositories (2008) is a highly recommended source for understanding how social and cultural factors can create a motivational gap between repository builders and faculty members:

The library community has built a solution [institutional repositories] based on a perceived problem (scholarly communication crises); but because the academics do not

perceive a problem that needs to be fixed, they are reluctant to adopt practices and policies imposed on them by others in the institution (Rieger 2008, under section 2).

Rieger concludes that most of the impediments to growing self-archived deposits in institutional repositories relate directly to the persistence of this gap. To make progress, institutional repository managers need to redouble their efforts to understand each discipline's value systems, research contexts and work practices. In this regard, library liaison outreach programs and work practice studies can guide repository developments and promotional programs.

### ***Reducing time and effort***

#### *Mediated deposit*

A comprehensive study suggests that two-thirds of self-archiving authors choose to do so using personal web pages, followed by research group or department web pages, then subject repositories, then institutional repositories (Kim 2010, 1914). A number of studies indicate that in the US and UK a large percentage, if not the majority of faculty deposits in institutional repositories are not truly *self-archived* but *mediated*, that is, undertaken by someone other than the creator of the work (Rieger 2012; Hanlon and Ramirez 2011; Armbruster 2011; Darby and others 2009). While mediated deposit may help to grow institutional repositories with content from faculty, taking on the task of mediating deposits is not without cost; offering a mediated deposit service has implications for institutional repository staffing. There is also the question of whether manually mediated deposit will successfully scale to a large number of deposits. In a study of time and costs associated with mediated deposit workflows in the Welsh Repository Network, Payne (2011) found a range between 12 and 15 minutes per deposit for workflows that did not include copyright clearance activities; her findings are roughly comparable to those of other studies she cites. For mediated deposit workflows that include copyright clearance activities, the time required per deposit would be greater, as described by Hanlon and Ramirez.

*Automated deposit*

Bulk deposits (large-scale uploads of repository content) and automated identification of objects for deposit are other techniques for reducing the time and effort needed to populate institutional repositories, for example as described by Shreeves (2009). Duranceau and Krieger's paper (2013) describes many implementations of automated deposit at different libraries, including some that use the new protocol SWORD (Simple Web-service Offering Repository Deposit). SWORD and SPI (Simple Publishing Interface) appear to be the way forward for making it easier to disseminate and re-use content and metadata in multiple systems and applications.

**“Google has won”**

A characteristic remark among the digital library experts who were interviewed for this book was something like “Google has won the discovery game.” History and the evidence presented in this book support this perspective. Looking back, Sergey Brin and Larry Page founded Google in September 1998. It is highly unlikely that digital library pioneers at that time imagined how the Google search engine would transform where and how scholarly communities look for and get research-quality information.

The Ithaka longitudinal study of faculty indicates that by 2009, 70% of faculty were often or occasionally using a common search engine to locate journal articles (Schonfeld and Housewright 2010, figure 5). A recent Ithaka study of how academic historians go about their work found that their day to day research practices have been fundamentally changed by technology and the web, and that Google and Google Books have become singularly important to them (Rutner and Schonfeld 2012, 18-19):

Interviewees use general Google searches to start the discovery process. For many of them, Google is the primary search tool in identifying archives that hold relevant materials ...Google is recognized as a tool that has expanded the

breadth of types of materials that an historian can access on a given topic, and introduce a researcher to collections that they were not aware of, even after years of working within a sub-field ... Interviewees widely acknowledged Google Books ...nearly all of them mentioned using it in some capacity...

### **SEO and ASEO**

The librarians and archivists who began digitizing cultural heritage materials in the 1990s could not have predicted the eventual centrality of Google or Google Books to humanists' research practices. Similarly, the earliest repository managers would not have expected the majority of repository traffic to come from Google or Google Scholar, and they would likely have been surprised by Arlitsch and O'Brien's advice (2012) to optimize their metadata for academic search engine crawling (see the discussion in chapter 4). Beel, Gipp and Wilde (2010, under section 2.1) describe ASEO (academic search engine optimization), which they define as "the creation, publication, and modification of scholarly literature in a way that makes it easier for academic search engines to both crawl it and index it." While Google Scholar is the undisputed leader among academic search engines at this time, Beel's article also identifies IEEE Explore, PubMed and SciPlore.org as academic search engines. Chapter 5 discusses the role of such sites as referrers to hybrid library collections.

Actively working with search engine optimization (SEO) techniques can be a controversial subject among librarians, as discussed by Onaifo and Rasmussen (2013). At the same time, these authors found in their study of public library websites that SEO techniques do positively affect library website rankings and the degree to which they attract users. Arlitsch and O'Brien were able to significantly boost the findability and use of their institutional repository content through ASEO techniques. Chapter 10 returns to the discussion of SEO's role for digital libraries.

### ***Web traffic analytics***

If a key objective of repositories is to maximize access to research-quality content, repository managers need to not only collect data about the nature of their repository's online traffic, but also understand and emphasize SEO approaches to improving visibility and use. Finding time for this type of work can inform strategies for integrating repositories in real-world research, teaching and learning practices.

While Google and Google Scholar are important referrers of searches to all repositories, the most successful repositories can be said to be destination sites as well (sites that people visit by going directly to their URLs through typing or bookmarks). At least with respect to discovery, the many remaining repositories function less as destination sites and more as content delivery mechanisms, responding to searches that originated elsewhere. The implications for repository interface design and development are discussed in a later section.

### ***Making institutional repositories more valuable***

This and other chapters have traced a variety of issues that can negatively affect the future prospects of institutional repositories. Table 8.2 brings together these barriers and offers possible service responses. While the table aggregates the points made across several chapters, it is important to understand that not all barriers and possible service responses apply to all audiences. Some service responses can be inappropriate in some contexts.

### ***Understanding and identifying repository audiences***

The starting point for using the table is to understand the specific needs of the audience(s) that a specific repository might serve. The next step is the purposeful selection of which audience(s) to serve. The selection of audience(s), combined with an understanding of the needs of that

audience or audiences, will reveal the most likely barriers and offer some possible service responses. No single repository enhancement is as important as deciding what audiences to serve and designing the repository with these audiences' needs and behaviors in mind. The aim is to establish a firm foundation for the repository within institution-specific communities of research and practice.



**Table 8.2 Barriers and possible service responses for institutional repositories**

Barrier	Possible Service Response
<p><b>Lack of clarity around repository’s purpose and focus</b></p>	<ul style="list-style-type: none"> <li>• Conduct needs assessment(s) of audience(s) to be served</li> <li>• Involve intended audience(s) directly in setting mission/purpose and goals</li> <li>• Find champion(s) in audience(s) to be served</li> <li>• Validate assumptions about intended audience(s) needs, content, expectations</li> <li>• Agree upon shared mission</li> </ul>
<p><b>Lack of awareness</b></p>	<ul style="list-style-type: none"> <li>• Conduct needs assessment(s) of audience(s) to be served</li> <li>• Articulate audience-specific unique advantages (value propositions)</li> <li>• Communicate value through branding</li> <li>• Plan and carry out audience-specific communications programs</li> <li>• Build relationships: reach out directly to audience(s), e.g., through liaison librarians</li> <li>• Increase discoverability of content on the web</li> </ul>
<p><b>Lack of participation and engagement (low deposit rates)</b></p>	<ul style="list-style-type: none"> <li>• Conduct needs assessment(s) of audience(s) to be served</li> <li>• Articulate audience-specific unique advantages (value propositions)</li> <li>• Identify a problem an audience(s) wants solved, and solve it</li> <li>• Lower barriers to participation; easy to get started</li> <li>• Find partners in intended audience(s): undertake collaborations with them</li> <li>• Welcome new types of content</li> <li>• Engage audience(s) with content, e.g., crowdsourcing, social sharing</li> <li>• Provide incentives</li> <li>• Provide usage evidence and statistics</li> <li>• Offer mediated deposit (or deposit to multiple repositories)</li> <li>• Undertake automated deposit</li> <li>• Promote and support deposit mandates</li> </ul>

Barrier	Possible Service Response
<p><b>Poor alignment with community work practices</b></p>	<ul style="list-style-type: none"> <li>• Study work practices; undertake user-centered design</li> <li>• Provide authoring support tools (virtual research environments; scholars' workbench concepts; version management; security/permissions)</li> <li>• Offer workspace for various types of content</li> <li>• Offer mediated and/or automated deposit alternatives</li> </ul>
<p><b>Complex legal framework</b></p>	<ul style="list-style-type: none"> <li>• Offer copyright advisory services and training</li> <li>• Provide information related to publishing and publication</li> <li>• Training</li> </ul>
<p><b>Traditions of scholarship</b></p>	<ul style="list-style-type: none"> <li>• Understand discipline-specific norms, peer review, reward systems, attitudes</li> <li>• Target efforts to those who are willing</li> <li>• Offer author information services</li> <li>• User-centered design: support disciplines' established workflows for depositing and credentialing new scholarly content; support collaboration with peers</li> <li>• Build relationships: reach out directly to audience(s), e.g, through liaison librarians</li> </ul>
<p><b>Discoverability</b></p>	<ul style="list-style-type: none"> <li>• Optimize discovery possibilities through ASEO and SEO best practices</li> <li>• Establish links from high-traffic sites (e.g., learning management systems)</li> <li>• Provide stable identifiers and URLs</li> <li>• Participate in registries and interoperability frameworks (e.g., repositories of repositories)</li> </ul>
<p><b>Sustainability</b></p>	<ul style="list-style-type: none"> <li>• Undertake business planning and establish metrics for success</li> <li>• Select one or more approaches to sustainability: institutional hosting/subsidies, sponsorship, partnerships, in-kind contributions, volunteers, etc.</li> </ul>
<p><b>Preservation</b></p>	<ul style="list-style-type: none"> <li>• Integrate digital preservation policies, systems and practices</li> <li>• Seek out partners</li> <li>• Participate in national or regional initiatives</li> <li>• Secure the right to preserve</li> </ul>

### ***Implications for interface design and repository development***

Selecting repository audience(s) and needs to be met has implications for interface design and overall development of the repository. To what extent will the repository serve as a content delivery or landing site for searches referred from elsewhere? Is the repository a destination site and if so, for which audience(s)? What local workflows will the interface support? Will the repository interoperate or provide content to other systems? These questions can only be answered in the context of how the repository is intended to support target audiences and specific needs.

### **Future of repositories**

This and prior chapters have explored a number of questions related to where repositories will go from here. Two important questions have not yet been discussed:

1. In what ways will repositories support digital data management and 21<sup>st</sup> century scholarly research infrastructure?
2. Today's thousands of repositories are more like a conglomerate (entities that can be conceptually grouped together but which remain distinct) than an ecosystem (a community of interconnected parts in an identifiable framework). To what extent are repositories likely to evolve into a sustainable, global ecosystem for capturing, making accessible, and preserving the scholarly record?

Brief discussions of these two questions conclude this chapter.

### ***Cyberinfrastructure, e-research and data curation***

#### *Background*

Since 2004, the NSF has issued a number of calls for grant proposals for research related to cyberinfrastructure and the stewardship of digital data (for details see Atkins et al. 2003). The

term “cyberinfrastructure” was first used by NSF and it refers to computing and data systems, repositories, instruments and high speed networks that together frame scholarly research environments and enable advanced capabilities. In the UK, Hey and Trefethen (2003) previewed the anticipated “data deluge” from e-science (computationally-intensive science carried out through internet-enabled global collaborations). They argued for digital data libraries and curation (activities including selection, organization, preservation, maintenance and archiving) for scientific data. Later the e-science prefix “e” was used to describe the same kinds of activities in the social sciences, humanities and so on. Current usage favors the more inclusive term “e-research” (distributed, data- and information intensive, collaborative research).

In her book on digital scholarship, Borgman (2007, xvii) stresses the point that “data have become an important form of research capital.” Data sets supplement traditional scholarly publications, and they serve as important outputs of research and inputs to new research. At present it is not clear how data and scholarly publications can be linked together in scholarly information infrastructures. Nevertheless a growing number of developments have encouraged more systematic e-research data sharing, management and preservation.

#### *E-research data sharing*

Important policy changes have occurred for those submitting and working on research grants from public funds. In 2010, NSF announced a new requirement to incorporate data management plans (National Science Foundation 2010) in grant documents. This development affirms the growing consensus reflected in the *OECD Principles and Guidelines for Access to Research Data from Public Funding* (OECD 2007) that open access to research data is critical to the progress of science and optimizes the results of investing public funds in research projects.

In addition, a number of developments and activities to document and enable good practice in digital data collection, citation, federation and sharing have occurred. This section mentions only a few of many. For example, DataOne (Michener et al. 2011) federates content from distributed repositories called member nodes, using an innovative scheme for using persistent identifiers (EZIDs) to manage and track digital objects across repositories. Crosas (2011) describes work at Harvard on the DataVerse Network (DVN). DVN supports individual researchers' data ownership and control, while at the same time enabling and incentivizing data sharing through a repository infrastructure for data identification, management, preservation, re-use, discovery and visibility. Brase (2009) discusses DataCite, an association that promotes and facilitates the use of persistent identifiers for datasets. Simons (2012) discusses a partnership with DataCite to mint DOIs for research data and provide them in a "Cite My Data" service.

The literature around digital data management is large. Anna Gold's two-part "cyberinfrastructure primer" provides a place to start (2007a; 2007b).

#### *E-research data and repositories in Australia*

A considerable amount of activity in Australia has explored the role of repositories for supporting data curation, for example Treloar, Groenwegen and Harboe-Ree (2007) at Monash University. Monash was a player in the development of the now-completed ARROW project (federating institutional repositories in Australia; not the same as the ARROW knowledgebase supporting rights management for Europeana). DART and ARCHER are two additional projects on researcher workflow and data management.

Wolski, Richardson and Rebello (2011) describe an initiative of Australian universities to develop a framework for feeding data into both a national research data service and university library discovery tools. Their infrastructure federates content created locally and makes it

accessible for sharing in different systems, supporting different discovery environments. Two of these are the Australian Research Data Commons, which supports discovery in Research Data Australia (RDA), and LibrarySearch, which supports discovery of hybrid library resources (see chapter 5 under 'A new kind of library catalog'). The work being done in Australia is encouraging to those who have hopes for a unified, large-scale service framework to interconnect e-research services, at least at the national level.

### *E-research, libraries and repositories*

The conversation about e-research, digital data and data curation that began in 2003 has been taken up by both information scientists and academic librarians. In 2005 ARL obtained a grant from the NSF Office of Cyberinfrastructure (National Science Foundation 2006) to convene experts and study new collaborative relationships between academic libraries and digital data producers. Their report (Friedlander and Adler 2007) lays out the issues and proposes roles for libraries in e-science and data management that have been the subject of further collaborations and studies. A Research Information Network report in the UK (CIBER 2010) describes the state of e-research support in four universities in the UK and concludes with recommendations, including a call for exploring expanded roles for information specialists to work with research teams. Much more information may be found in Bailey's bibliography on academic libraries and e-science (2011).

ARL reported the results of a survey of how US members are providing infrastructure or support services for research data and e-science (Soehner, Steeves, and Ward 2010). In an early report of work on a survey, Corral, Kennan and Afzal (2013) investigated the bibliometrics and e-research data support services of 140 libraries in New Zealand, Australia, the UK and Ireland; a majority anticipated involvement in some aspect of e-research data support, with the primary areas of focus being technology infrastructure, data deposit services and policy development.

Corrall and her colleagues found that a major constraint on the development of e-research data services in libraries is the gaps in librarians' data management knowledge, skills and confidence. Others have also recognized this gap; Borgman's syllabus (2012) for her new course on data and data curation points to key topics and recommended readings for educating oneself in this emerging specialization for librarians, information scientists and digital library specialists.

While there are many encouraging signs that libraries and the repositories they host will have a role in supporting e-research and data curation, it is too soon to predict the outcome of current initiatives. One of the experts interviewed for this book cautioned:

We really don't understand data management or digital libraries of data. Sharing, managing and making data sets more generally discoverable, accessible and reusable are very difficult problems requiring socio-technical advances and structural changes. Developing solutions will take a very long time.

It would be unfortunate for digital library history to repeat itself, with information scientists and librarians pursuing parallel but essentially separate paths. Looking at what has transpired so far, there is some troubling evidence that this could happen. More encouraging is the large-scale, NSF-funded project at Johns Hopkins, which is focused on building a technical framework on which an institution's data management and curation services can be layered (Mayernik et al. 2012). It is also encouraging that some groups are collaborating with the scholars for whom they are envisioning services (see Walters 2009, describing a project at Georgia Tech).

The extensive literature review offered by Corrall, Kennan and Afzal (2013) also reveals some work to assess needs and develop campus partnerships. Many current initiatives are driven

from worthy motives—but it must be admitted that few seem to be growing organically out of the needs and preferences of one or more specific scholarly audiences. A personas-based case study suggests that researcher receptivity to library-managed data curation is discipline-driven (Lage, Losoff and Manness 2011). In keeping with lessons learned about institutional repositories, it appears that the success of university- or library-hosted data repositories will depend on understanding and engaging with specific communities of scholars in specific scholarly disciplines.

***Future repositories: ecosystem or conglomerate?***

The final section of this chapter explores the prospects for repositories' becoming key components of larger, more purposively coordinated research information environments of the future. From their recent survey of academic libraries with repositories, Burns, Lana and Budd (2013) concluded:

No one can predict the future of institutional repositories at this time and it remains to be seen if individual institutional management of a repository is the most efficient and effective means of operation. A question that should be asked of the users of repositories is whether their needs are met by the dispersed model of repositories that exists at the present time...

Burns' question is not new. Herbert Van de Sompel made the case some years ago for replacing the current scholarly communication system with “an innately digital scholarly communication system that is able to capture the digital scholarly record, make it accessible, and preserve it over time” (Van de Sompel et al. 2004). He proposed a “fundamental re-engineering to a network-based system that ... provides interoperability across participating nodes.”



### *Ecosystem*

Van de Sompel's 2004 call for a network-based system of loosely-coupled, communicating services has been frequently cited. It led to a new set of web-compatible standards for aggregating resources (OAI-ORE) and a "remix and reuse culture" for sharing scholarly content (see for example Lagoze et al. 2010). Others around the world have been wrestling with the complex problem of better integrating heterogeneous content into dynamic research information environments. Starting in 2005, a number of JISC-supported reports appeared referring to repositories as elements of an emerging "ecosystem" or "ecology" of scholarship, knowledge creation, discovery, use and transfer. Heery and Anderson (2005) called for an "ecology of repositories" framed by common standards, protocols and interfaces that would support "a distributed network of repositories interacting with national and international initiatives." The following year, Heery and Powell (2006) provided a roadmap for the development of such a framework.

Various supporting mechanisms to unperpin an ecosystem of repositories or digital libraries are present or emerging. OAI was an early breakthrough enabling multiple repositories to be federated in aggregations like OAIster or DAREnet (see chapter 4). OAI's evolution and the introduction of ORE have already been mentioned. In addition:

- DRIVER has demonstrated the feasibility of a common infrastructure for networking European digital repositories and federating their contents (Manghi et al. 2010; Peters and Lossau 2011)
- National-level programs such as that of the Dutch national library have specified cross-domain aggregation and interoperability standards, enabling the wide discovery, use and reuse of Dutch digital content, including in Europeana (Janssen 2011)

- The authors of the DELOS Reference Model and the DELOS Manifesto, now active in the EU co-funded D4Science-II project exploring e-infrastructures and Virtual Research Environments, have introduced the concept of a Knowledge Ecosystem model, which enables interoperability among multiple independent e-infrastructures (Candela, Castelli and Pagano 2011, 14-15).
- Sefton and Dickinson (2011), describe how repositories are or could be a key component in an “Academic Working Environment” or AWE, an evolving ecosystem supporting researcher workflows and services in Australia.

### *Conglomerate*

Both “ecosystem” and “ecology” imply an environment, albeit complex and dynamic, with known or discoverable interrelationships and systematic interactions among the elements of the system. By contrast, a “conglomerate” brings component parts together into a whole, but in and of themselves, the parts have no relationships and they remain distinct. Despite the progress toward a more purposively coordinated approach exemplified by the previous examples and advocated by so many in the digital library field, at present the world’s repositories are for the most part isolated and dispersed. Taken as a group, they can be said to be a conglomerate—a conglomerate with a level of cross-repository search thanks to the effects of Google Scholar, other academic search engines, and search engine crawling, which brings much repository content together for global discovery and use.

While some of the digital library experts interviewed for this book are optimistic that something like an organized ecology of worldwide repositories will emerge, others expressed doubts whether a well-integrated, communicating set of services and systems is a likely scenario. The problems and issues they mentioned included:

1. The feasibility of large-scale collaboration among numerous and far-flung organizations; one interviewee commented “we overestimate the usefulness of collaboration and underestimate the challenges.”
2. The difficulty of achieving sustainability; one interviewee wondered “who would pay for an interconnected system?” There is a mismatch between the benefits of an interconnected ecosystem to scholarly endeavors, which are global, and the means by which repositories are commonly funded, which are usually local, sometimes consortial or national, and rarely international.
3. The likelihood of many diverse repository-hosting organizations’ adopting digital library-specific standards requiring additional effort. One interviewee noted “web companies have done well because they didn’t ask contributors to do anything.”
4. The difficulty of achieving repository interoperability as originally conceived (i.e., through digital library standards-based cross-searching of many diverse collections at once). One interviewee noted “it is easier to replicate content [in multiple repositories] than to integrate it with distributed search.”
5. The complexities of the legal framework governing access to scholarly content. One interviewee noted “open access is not the silver bullet for fixing scholarly communications that some claim it can be; a system with ‘shades of access’ is more realistic.” In addition to the legal issues, managing authentication and authorization adds technical complexity.
6. The importance of discipline-specific solutions that arise from the community they are intended to serve; a generic solution around an organized ecology of repositories would represent a significant break with scholars’ preferences for working within their own fields and in keeping with extant scholarly reward systems.
7. The fact that deposit in institutional repositories is at present haphazard; the lack of scholars’ commitment to and engagement with most repositories.

8. The disruptive, rapidly changing socio-technical environment of which scholarly repositories are a part. It is difficult to know what will be the next innovation to break down former barriers and open up formerly inconceivable options. The prospects of the semantic web and linked data to create new solutions cannot be predicted at this point.

### *Force field analysis*

Figure 8-1 is a force field analysis—a framework for evaluating the factors that could drive or restrain the emergence of a sustainable, global ecosystem of scholarly repositories. Driving forces are on the left, and restraining forces are on the right. The previous section discusses the restraining forces. Earlier sections have discussed the driving forces; in summary they include:

1. The promise of a better, more economically and socially valuable system supporting scholarly communication that is global in scope and that supports the free flow of ideas, individual empowerment, teaching, learning and the advancement of knowledge now and for future generations
2. The existence of successful regional, national and international programs
3. The existence of *de facto* cross-repository search through common and academic search engines, Google Scholar (and/or its descendants)
4. Advocacy for maximizing the impact and utility of funded research, and the emergence of policies supporting open access to publicly funded research around the world
5. Advocacy for bridging digital divides not only in developing countries, but also within countries where only some have ready access to licensed scholarly content
6. Digital library practitioners' evolution toward building on and for the web, using low-barrier, lightweight standards, and away from more complex library/repository-specific standards; and new approaches for facilitating deposit and content reuse/exchange such as the SWORD protocol

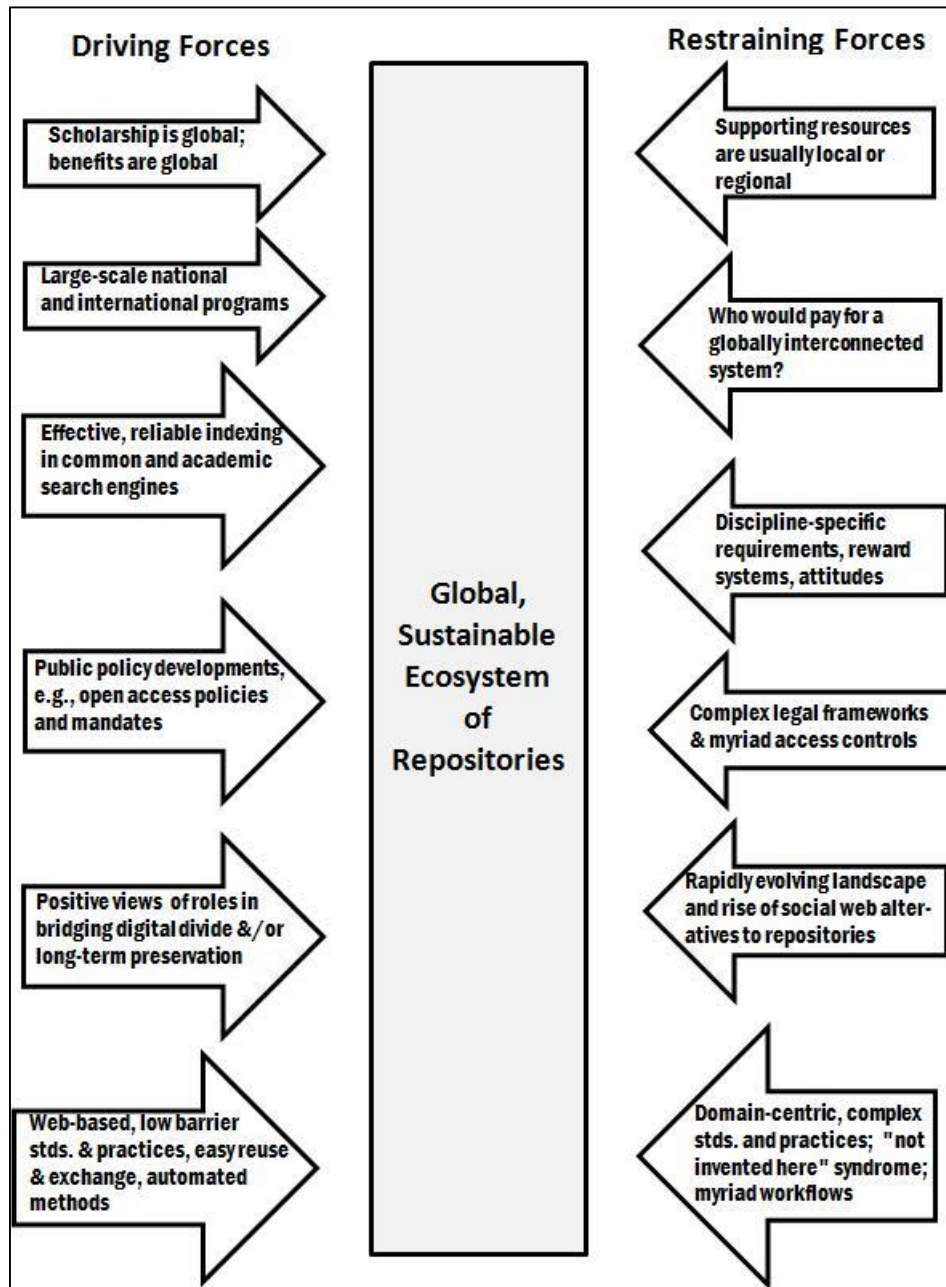


Figure 8.1. Force field analysis of the emergence of a sustainable, global ecosystem of repositories

## **Conclusion**

Researchers disseminate their findings to make an impact on their field and for personal career progression. Both objectives are served by publishing in the peer-reviewed journals with the highest impact factors. Those repositories with a firm foundation in the scholarly communities that they serve have provided a needed service for sharing early results and establishing claims to new discoveries. Others, including most institutional repositories, have not yet achieved an integral role with scholars. On the other hand, taken as a whole, repositories have already contributed substantially to the achievement of a digital library social role—that of broadening access to the results of scholarship for a large number of people.

Depending how they evolve from this point, repositories have the potential to support other digital library social roles as well: enriching teaching and learning; providing enabling infrastructure; enabling multidisciplinary knowledge work and knowledge transfer; increasing scholarly productivity; preserving and curating intellectual assets for future use; and more. Overall the prospects of repositories are favorable, but much work needs to be done. The challenges mentioned in this chapter include:

- Better understanding the needs and work practices of the communities to be served
- Better understanding discipline-specific norms, peer review, reward systems and attitudes
- Improving awareness and recognition; better articulating the value of repositories in ways that resonate and align with the needs of a variety of audiences
- Increasing deposits
- Assisting scholars with the complexities of open access and rights issues
- Improving the productivity of scholars both individually and collectively

- Aligning repository practices to optimize discoverability in search engines and academic search engines
- Preserving and curating content; ensuring long term access to intellectual assets
- Achieving sustainability; undertaking business planning; establishing and achieving metrics for success

To the degree that future repositories integrate well into the web and use web-based approaches, blend well into how their communities do things, attract attention to all kinds of collections, and provide the means to easily disclose, reuse and share content, they can be successful. Many of the pieces for assembling repositories into a useful global configuration are in place or emerging. Whether the pieces will in fact converge to produce a revitalized framework supporting scholarly communication is hard to say, but the potential exists. Much may be determined by the impending impact of the emergent (and presently chaotic) information space defined by e-research initiatives, scholarly social networks, repositories built around researchers (rather than their works), shifts in scholarly reward systems based on new metrics and the emergence of semantic web applications. This new information space is the subject of the next chapter.