

Innkeeper at the Roach Motel

Dorothea Salo

Digital Repository Librarian

University of Wisconsin Digital Collections Center

Abstract

Trapped by faculty apathy and library uncertainty, institutional repositories face a crossroads: adapt or die. The “build it and they will come” proposition has been decisively proven wrong. Citation advantages and preservation have not attracted faculty participants, though current-generation software and services offer faculty little else. Academic librarianship has not supported repositories or their managers. Most libraries consistently under-resource and understaff repositories, further worsening the participation gap. Software and services are wildly out of touch with faculty needs and the realities of repository management. These problems are not insoluble, but they demand serious reconsideration of repository missions, goals, and means.

Introduction

“[The institutional repository] is like a roach motel. Data goes in, but it doesn’t come out.” (Dorothea Salo)

Institutional repositories have not fulfilled their early promise of increased access to the scholarly journal literature through faculty initiative. Academic libraries twist in the wind, enticed by possibilities but disillusioned by results, uncertain how to proceed. Thus far, at least in the United States, doubts about the viability of institutional repositories have been kept quiet or denied altogether. As long as libraries and repository managers remain silent about the current deplorable situation, however, no one can rectify it. Even in the current difficult environment, institutional repositories and their managers can find plenty of work to do, given realistic goals, support from colleagues and administrators, and software that serves real needs rather than hopeful ideologies.

Pseudonymously-sourced quotations that begin each section of this article have been stripped of identifying detail, but within the limits of faulty memory they are genuine and accurate. As both the repositories I have run have used DSpace, examples in this

article may be too frequently drawn from DSpace and its developer and user community, for which I can only apologize.

Faculty and self-archiving

“Institutional repository? Forgive me, but – that sounds vaguely obscene.” (Graduate student in psychology)

“What? No! I’d never want those [preprints] on the web! They’re not authoritative! I’d never use them, either!” (Senior professor of engineering)

“[Engineering faculty] don’t even know the library exists. They never go there; they download all they need. The library doesn’t even register with them.” (Engineering IT manager)

Except in a few disciplines with thriving disciplinary repositories and in the bare handful of institutions that have imposed deposit mandates or are blessed with an extraordinarily engaged faculty, unmediated faculty-initiated self-archiving has failed abjectly thus far. “At a median growth rate of 1 item a day, IRs in America will likely not achieve the critical mass to significantly impact open access or change modes of scholarly communication for some time to come.” (McDowell 2007) This calls the entire existence of institutional repositories into question, predicated as they were on the assumption that faculty would deposit, describe, and manage their own material.

Open access alone is wholly insufficient as a selling point; most faculty have yet to assimilate the concept, and initial reactions are as likely to be negative as positive (Swan 2006). Consider the philosophy professor who had an online bibliography plagiarized and therefore “cannot call [his] experiment in open-access publishing a success.” (Vincent 2007) However faulty his assessment of his experience, his reality informed faculty reactions to open access all over the country, and he is far from alone in his skepticism. In my experience, faculty who believe that self-archiving may threaten their rights over their work, their relationship with their favorite publishers, or their status in their disciplinary communities, will not consider it; faculty who believe open access threatens those publishers or those communities are vocal open-access detractors.

Having little else to offer, repository managers have resorted to touting unbreakable URLs and preservation capacity. While these are genuine benefits, faculty are only beginning to become aware of digital curation, feel little if any personal connection to the issue, and so cannot easily be sold on it. Preservation of the fruits of faculty labor has never been a faculty problem; it is the problem of libraries and (secondarily)

publishers. Even faculty aware of the problems do not believe that librarians can help solve them: “an inherent culture of self-sufficiency in the generation and organisation of data militates against what might be viewed as prescriptive intervention by knowledge management professionals” (Pryor 2007). An examination by JISC agrees: “There are perceived risks [to repository deposit, according to researchers], since most IRs are managed by librarians without deep subject knowledge, rather than science experts and there is a likelihood of technical problems with regard to formats and metadata management.” (Lyon 2007). This self-reliance can turn pathological. I have heard credible, often firsthand stories of datasets kept in a principal investigator’s home-basement data-center, and of immense photographic-image collections wasting away to uselessness undescribed and unarchived, because principal investigators hired graduate students in the discipline with a little programming expertise but no knowledge of information management to fix the problem. Some researchers, it seems, would rather turn *anywhere* than to a librarian!

Tenure and promotion loom large in the cautious stance of faculty toward repositories (University of California Office of Scholarly Communication 2007). Many researchers (and nearly all tenure committees) still regard digital materials with distrust and scorn even as their usage preferences shift dramatically toward them; repositories suffer from the same prestige gap as any other all-digital environment. Peer-review aside, prestige of publication venue plays the single largest role in faculty decisions about their research content (Hendler 2007); no other measure of quality or impact comes anywhere close. Even the oft-lauded open-access citation advantage (see Hitchcock 2007 for bibliography) flies in the face of standard faculty notions of meritocratic scholarly authority, unable to compete with the known, established career benefits of publication in prestige journals. In an ideal academia, an article would garner citations as a function of the prestige of its publication outlet and its own quality. Since self-archiving confers no prestige and makes no representation of content quality, it *should not* make any difference to an article’s readers and citations, at least in the mind of a researcher accustomed to the current scholarly-communication system. Younger scholars may be attracted to self-archiving as a way to game a prestige system otherwise stacked against them, but older scholars are liable to resist the very idea of an open-access citation advantage.

Disciplinary differences create challenges of their own. Several disciplines whose journals figure heaviest in the serials crisis rarely notice the library and its services, as they access nearly all their materials through online services the library provides transparently: engineering and related fields, computer science, hard sciences, and so on. Medical libraries are shrinking staff, convinced (however wrongly) that the medical library is an anachronism. Library-managed institutional repositories face particularly sharp struggles in these disciplines, as their researchers are often disinclined even to accept librarians as fellow professionals.

Patchy and frustratingly inconsistent self-archiving policies among publishers form a potent barrier to faculty participation, even as they allow publishers to protest virtuously that they are playing along with the open-access movement. Faculty (and many librarians) may not understand the subtle differences between preprints, postprints, and publisher PDFs; trying to educate them drives them away, their interest in self-archiving gone. Although arguably much of the material in question ends up in disciplinary repositories or researchers' professional or personal websites anyway, often illegally (Wren 2005), this helps institutional repositories not in the slightest.

Publishers have muddied the waters further with burdensome and heterogeneous requirements for acknowledgement of published versions; worse, repository-software rigidity has made compliance with those requirements technically impossible in some cases – DSpace, for example, cannot create a link from Dublin Core metadata except in fields designated solely for URIs, so no DSpace repository can comply fully with a set-phrase requirement, such as Springer's, that includes a link. As for publisher-policy discovery, coverage of the journal universe at the well-known SHERPA/ROMEIO database (<http://www.sherpa.ac.uk/romeo.php>) is uneven at best, and especially poor with independent and small-society journals. Worse yet, because SHERPA/ROMEIO has only recently begun to offer an API, repository managers have not yet automated the process of checking whether mediated deposits are legal, leaving permissions management a time-consuming manual process fraught with risk.

The first wave of promising technological enticements has generally come to naught. The University of Rochester's Researcher Pages modification to DSpace (Foster and Gibbons 2005) made a splash when it was first announced, but the code was not widely adopted elsewhere and eventually died a lingering death (more for technical reasons to be discussed later than because the idea was poor). An email-the-author button on

embargoed content in EPrints had to be removed briefly because it proved *too* popular (Harnad 2006); although this innovation became part of EPrints core code, it has not caused a sea-change in self-archiving rates. None of these interventions has successfully challenged the underlying faculty apathy that already vitiates marketing efforts; nor has it automated the process of capturing content already available on the Web in order to compensate for faculty failure to act.

Several vocal self-archiving advocates engage in a species of doublethink around researcher apathy toward self-archiving. “All researchers, without exception, do want to make their publications P-OA, and the online infrastructure for publication-archiving... already has all the requisite functionality for this,” reads the opening paragraph of an article ironically entitled “Incentivizing the Open Access Research Web” (Brody et al. 2007). Surely something all researchers want needs no further incentives? The desire to paint the picture in brighter colors than warranted is understandable, but counterproductive. University administrators who hear that all researchers already want open access will not themselves support or defend it, especially with scarce budgetary and staff resources. Library administrators who hear that all researchers already want open access will go to their repository managers in perfect faith asking why their repository’s capture rate is so dismally low, and what response can a repository manager possibly muster?

The failure of self-archiving ideology to win faculty hearts and minds created piercing calls for institutional self-archiving mandates, from Stevan Harnad’s repeated assertions that “the only swift and sure way to break free of this paralytic cycle is for researchers’ own institutions... and funders... to mandate open access self-archiving” (Harnad 2006) to Arthur Sale’s more moderate suggestion of a “patchwork mandate” starting in specific departments to provide a base for later expansion to the entire institution (Sale 2006). No institutional-repository manager has openly opposed the idea of a self-archiving mandate. None would. Unfortunately, institutional-repository managers are mice in mouseholes, bells in hand, looking out helplessly at the immense faculty feline. They have no authority over faculty. They have little to no influence in either library or university administration. Since their access to departments is generally mediated through other librarians, even the patchwork mandate is beyond their ken, unless they have the rare good fortune to find an influential, self-archiving-savvy faculty member to smooth the way.

Institutional repositories as currently implemented are parasitic on existing research and scholarly-publishing processes. They do not facilitate prior peer review or colleague feedback, and the many interesting ideas surrounding open review, post-publication review, and overlay journals have yet to come to fruition. They do not appropriately manage or version files before they reach their final form, although this is a commonly-expressed faculty need. They do not help researchers share unfinished work with their collaborators or a select few reviewers. They do not help with grant applications, submissions to publishers, or visibility in existing disciplinary repositories. They do not help a researcher achieve tenure, promotion, or the next round of grant funding. What institutional repositories offer is not perceived to be useful, and what *is* perceived to be useful, institutional repositories do not offer.

In the absence of a compelling value proposition for faculty, repository-specific marketing efforts are futile, as every investigation into the subject to date has shown. Simply put, the institutional repository and services associated with it must provide value to faculty *on faculty terms* before it will see more than scant, grudging use. The institutional repository as currently imagined and implemented is a pig; no marketing initiative has sufficient lipstick to disguise it, and faculty will not initiate mandates requiring them to kiss it.

Libraries, librarianship, and IRs

“This is Dorothea Salo. She’s our – she does all kinds of nifty digital stuff.”
(Librarian)

“We don’t need to be running all that fancy digital stuff. We need to hire some *real librarians*.” (Librarian)

“Oh, institutional repositories? No, we’re teaching about repositories like [specialized toll-access vendor database].” (Library-school professor)

Academic librarianship has responded tepidly at best to open access in general and self-archiving in particular. Librarianship boasts two disciplinary repositories, E-LIS and DList, but both together contain under 10,000 items, barely a trickle of the articles, book chapters, presentations, and other materials generated by the discipline each year. Librarian awareness of self-archiving is patchy at best, and library professional organizations are not educating their rank-and-file members about it. Indeed, insofar as the American Library Association and Association for College and Research Libraries are themselves toll-access journal publishers, they have a conflict of interest that may be

preventing them from adopting the far more activist stance of the Association of Research Libraries, which does not publish journals.

Top library administrators are as a rule little better-informed than the rank-and-file about scholarly communication in general and open access in particular. Platitudes in the literature do not help them come to grips with the serious commitment and difficulty involved in running a repository. Typically, they leave repository managers starved of resources and do not undertake any outreach or education work themselves; this cripples repository efforts because only top library administrators can effect changes of attitude and behavior among institutional administrators such as provosts, deans, and department heads. "If top-level library administrators do not have a strong interest in and adequate understanding of scholarly communication issues as well as a real commitment to foster change," warns Charles W. Bailey Jr. (2007), "scholarly communication programs are hamstrung, and they become token efforts or die."

Truly shocking is the extent to which librarians' contributions to the open access movement have been ignored outside librarianship, and librarians themselves openly derided as contributing to problems rather than solutions. John Willinsky's landmark book *The Access Principle* (2005) barely mentions libraries except as funding sources. The useful essay collection *Open Access: Key Strategic, Technical, and Economic Aspects* (Jordan 2006) does not address the impact of open access on academic libraries or even include "libraries" or "librarians" as an index term, inviting the bizarre conclusion that libraries are *not* strategically, technically, or economically important to open access. "Maddeningly pedantic and out-of-touch at times" says self-proclaimed open-access archievangelist Stevan Harnad of librarians (Poynder 2006a); Richard Poynder wonders "Maybe it is also time to think the unthinkable, and walk away from the library as well" (Poynder 2006b).

Confusion therefore reigns inside libraries. Institutional repositories do not fit into librarianship's traditional quality and authority heuristics; not all content is the traditional peer-reviewed research article, and most of it is not available via the vendor-provided indexes librarians trust. Librarians are therefore less active than might be desired in using and promoting repositories and cross-repository search engines. Repository management is a new subspecialty, so new that most academic librarians have no idea even how to introduce repository managers to other librarians and (more importantly) to faculty. When asked to promote the repository themselves, most

librarians do not understand why they should or how to go about it, nor do collection developers and liaisons consider content recruitment for the repository part of their mandate. When asked to deposit their own content, librarians are as balky as faculty.

This combination of disregard and contumely only adds to the immense difficulty all repository managers have earning respect from faculty, fellow librarians, and library administration, both for the repository itself and for their role in running the repository service. Poor repository uptake among academic librarians invites faculty to charge their libraries with hypocrisy, particularly at institutions where librarians are tenured: if librarians themselves do not adopt the very practice they are so busily evangelizing to faculty, can it really be any good? Why should mandates cover faculty and not librarians? Yet no tenure-track libraries with institutional-repository programs have mandated deposit for their own staff's published and presented materials.

As repositories languish empty and opprobrium pours in from all sides, repository managers -- many of whom are staunch advocates of open access -- become demoralized and distance themselves from the open-access movement in general and institutional repositories in particular. Unclear -- and where they are clear, unmeetable -- performance standards for repository managers leave perpetual creeping worry that a repository manager will be an easy blame target for slow repository uptake, even when that manager has no direct control over faculty behavior, no resources, and no meaningful support from colleagues, library administrators, or the larger institution.

Repository managers have sadly *not* responded by forming a coherent community of practice. A Web bulletin-board service and mailing list I started in hopes of stimulating development of such a community folded within months for lack of uptake. Repositories have no defined publication outlets specific to them; research and case studies relating to institutional repositories are most often published in "special issues" of journals typically focused more broadly on library systems or scholarly communication; I was involved with an effort to start a journal on open access which had to be shut down for lack of submissions. To some extent, the repository-manager community is fragmented by software choices. A typical repository manager's first critical need is for technical support, which leads to dependence on tool-specific mailing lists. The dspace-general mailing list, for example, frequently holds discussions on matters of policy and procedure appropriate to all repository managers, not just those

using DSpace. Whatever the reasons, however, repository managers are isolated from each other and from the libraries that employ them.

Neither the open-access literature nor the library literature has gone much beyond threadbare platitudes in discussing how best to fund, staff, and assess institutional repositories, further confusing the question of their viability. This ignorance, added to the dearth of clear, attainable, measurable successes, produces pernicious effects. Repository managers understandably focus on filling the repository at all costs, since the easiest (though undoubtedly the least useful) measure of repository success is growth in collections (even empty ones) and items (even useless ones). This has led to repositories becoming an outright joke among university-press publishers: “institutional repositories so far tend to look like ‘attics’ (and often fairly empty ones), with random assortments of content of questionable importance” (Brown, Griffiths, and Rascoff 2007). It may also contribute to the expressed disdain for librarians among vocal open-access advocates, and the lack of comprehension among librarians why repositories collect student work, conference slideshows, and other non-traditional materials. Repositories and their managers fare even less well against better-considered metrics such as Les Carr’s measurements of steady deposit rates and growth in departmental involvement (Carr 2007). It is worthwhile, then, to examine how libraries actually *do* staff institutional repositories, because each option carries consequences for cost, service quality, and faculty responsiveness.

The maverick manager model

A librarian brought in specifically to run the repository is a “maverick manager.” Her job description usually includes policies and procedures development, outreach, training, metadata, maintenance chores such as batch imports, and permissions management; it may include programming, systems administration, or web design as well. As yet the maverick manager has no well-defined place in the library’s organizational structure; she may report to digital collections (in libraries fortunate enough to have such a unit), library systems, or special collections and archives.

The maverick manager has considerable freedom to experiment, but few to no resources. If she cannot herself program, or if she is walled off from systems administration as many librarians are, she cannot take advantage of useful plugins and

patches not directly incorporated into her software platform, nor can she do much to brand her repository or its contents, a stumbling block with many faculty. Since she has no budget, she cannot sponsor educational events, nor can she order educational materials from SPARC or go to the campus print shop for attractive, attention-grabbing print materials to distribute to faculty. Since she has no staff and no dedicated equipment, she generally cannot digitize analog faculty materials. This is a serious stumbling block at present, since nearly all faculty now straddle the analog-digital divide; faculty who wish all their work made available online immediately turn away from a service that only helps them with born-digital materials. Since she is only one person amidst hundreds or thousands of faculty, her ability to search out content and perform mediated deposits on behalf of faculty is limited, and cannot scale to the entire campus.

The maverick manager has no institutional power; she can wheedle but no more. Since as a new hire she is generally new to the campus, she does not have an institutional network to help her, either. Because she is one person and faculty are legion, nearly all her contact with faculty must be mediated through other librarians, either liaisons or administrators, any or all of whom may stonewall her or simply not care or know enough to assist. She cannot mandate self-archiving even in her own library, making the strident calls from open-access luminaries for self-archiving mandates a wistful dream at best. Without lures, and without administrative backing both inside the library and at high levels of the institution, the maverick manager cannot possibly attract the influential support she needs among faculty to achieve even patchwork mandates.

The maverick manager cannot unilaterally start most proven repository-population strategies, such as electronic thesis and dissertation programs; such is the culture of academia that she does not have the stature even to suggest them. Nor can she count on the assistance of her librarian colleagues with mediated deposits. Libraries that started institutional repositories on the assumption that faculty would perform and manage their own deposits did not plan for workflows in which librarians push the buttons at the behest of faculty. The maverick manager's colleagues are therefore not trained on repository deposit, not told they should canvass for material and be prepared to handle deposits and licensing, and (crucially) not given time or encouragement to perform repository-related tasks. The maverick manager is entirely on her own.

Willing, capable maverick managers appear to be few, and either eminently promotable, difficult to retain, or both. George Mason University (whose repository I managed for eighteen months), the University of Wisconsin (whose repository I currently manage), and North Carolina State University have each had to advertise for a repository manager at least twice in three years. Given the morale and career-development problems inherent in running a program that not only does not but *cannot* achieve its stated goals without a great deal more institutional commitment than currently exists, this state of affairs should not be surprising.

The “no accountability” model

Some libraries have asked information-technology units to build an institutional repository, or contracted with an outside vendor to host one, while dispersing responsibility for promoting and populating the IR among existing librarians, usually liaisons or collection developers. Too often, this dispersal of responsibility is not accompanied by clear goal-setting, performance expectations, appropriate time allocation, or administrative support. Existing staff have multiple competing claims on their time already, and the learning curve for repository software, for copyright questions, and for general open-access explication is more substantial than usually admitted. Since library administrators assume that faculty will perform most repository-related tasks, no provision whatever is made for mediated deposit or digitization. Unsurprisingly, this approach has yet to prove effective in filling the repository or promoting open access on campus. The repository becomes the library’s redheaded stepchild, its local unimportance only exacerbated by the general ignorance and apathy surrounding open access in academic librarianship.

The consortial model

To spread technology costs, some libraries are offering repositories consortially. The classic example is the Washington Research Library Consortium’s Aladin Research Commons (<http://aladinrc.wrlc.org/>), managed by WRLC’s technical staff and a committee consisting of librarians from each participating institution.

While this model indubitably offers efficiencies in systems administration and hardware purchasing, it exacerbates difficulties with repository outreach, technical

support, and content development. Consortium employees nominally in charge of the repository have almost *no* contact with faculty at the institutions forming the consortium. Marketing and outreach therefore devolve onto the same multi-institutional committee concerned with repository policy; unfortunately, repository policy and technical issues have a way of devouring meeting time, leaving outreach strategy an afterthought. Mediated deposit is generally out of reach, as neither the member institutions nor the consortium takes responsibility for funding and staffing it. Committee members rarely have much influence at their home institutions – if anything, even less than the maverick manager, whose institution has at least committed to funding her job! -- and the committee itself is powerless to shape library or faculty policy at individual institutions, so the consortial repository is completely powerless to effect change.

Because of the abysmal design of most repository software, many sorts of basic assistance (such as resolving depositor errors, handling batch imports, or opening a new collection) must be handled by technical staff at the consortium. A faculty member who has a technical issue with a consortial repository must therefore go through as many as four layers of staff to resolve it: a librarian at the institution who is not directly involved with the repository reports the issue to the institution's point-person for the repository, who in turn reports the issue to the consortium's repository manager, who finally reports it to consortium technical staff to be fixed. What outstandingly dedicated faculty member will report a problem twice?

A perhaps more worrisome problem is the uncertainty of consortial funding for a service that tends to receive uneven use by individual members of the consortium. One intransigent campus with influence can scuttle a consortial repository; alternately, the consortium as a whole may set short-term budget priorities that do not favor projects with long-term or uncertain payoffs.

The cooperative model

A very few institutions have formed a strong nucleus of interest and effort around their repositories. Faculty rather than librarians initiated the efforts leading to Ohio State University's KnowledgeBank (<http://kb.osu.edu/>), and since the project's inception,

faculty sense of ownership in it has meant steady deposit growth and considerable unmediated self-archiving (Melanie Schlosser, personal communication).

Other libraries have quietly accepted that mediated deposit is a necessity if an institutional repository is to be filled at all, and have built a cross-functional repository staff in furtherance of that goal. Not only do these librarians handle mediated deposits, they actively canvass institutional websites, disciplinary archives, and the open Web for content to deposit. Still others, such as the University of Illinois at Urbana-Champaign, hire dedicated developers to fill in the many gaps in the existing state of the repository—software art, or build automated workflows that pull content from tools and sites that faculty already employ.

Curiously, these repositories tend to hold their peace in the library literature. Perhaps they are too busy making their repositories successful to pontificate; perhaps they too would prefer an alternative to steady, time-consuming work. Their reticence, sadly, contributes to pluralistic ignorance among library administrators and repository managers of the real costs of running a well-functioning repository.

The downside of this model is its expense, chiefly in staff and development time; the enticing Ohio State situation is not presently achievable at most institutions owing to faculty apathy. Library administrators who believed that institutional repositories would run themselves must consider whether library priorities justify such expenditure. Nonetheless, some prickles of interest in institutional repositories can be seen in (for example) technical-services staff looking to expand their responsibilities beyond MARC cataloguing: “There exist numerous known and countless unknown digital objects on every academic campus whose current value and future sustainability could be substantially increased through aggregation in a trusted repository. Technical services units are well positioned to lead this work.” (Medeiros 2007) Such interest may form the core of an effective repository team.

Institutional repository software

“S[ocial] S[cience] R[esearch] N[etwork] gives them download statistics by author – they use that in tenure packages.” (Librarian)

“Where can I get support? Is there a vendor?” (Potential repository-software adopter)

“Throughout the life of the dataset, different things are added as annotations, as you analyze the data.” (Microscopy analyst)

“Knowledge-management solutions will be judged by the strength and flexibility of their security.” (Engineering IT manager)

The software underlying institutional repositories is quite young; EPrints was first released in 2000, DSpace in 2002, and Fedora in 2003. Hosted vendor offerings in this market are even younger. It is no surprise, then, that the software and services available to repository managers are overengineered yet underdesigned. Unfortunately, the current state of the repository-software art is so poorly managed, and so heavily flavored by unworkable ideology surrounding unmediated voluntary self-archiving, that improved offerings will take some time to appear.

Repository software serves open-access ideology decently well, on the whole. All three open-source packages have fairly solid back-ends from a preservation perspective, though compliance with the OAIS model is still lacking in DSpace and EPrints. All three manage discrete research articles quite well. All three (with reservations in the case of Fedora) provide a simple ingest interface for faculty to self-archive their own materials, the wide and toll-free dissemination of which is the chief driver behind the open-access movement.

The leading software and services have not traveled an easy road in libraries, however. All three open-source packages offer varying quantities of installation and maintenance headaches, expensive hardware demands, customization and development hassles, and poor fit with existing library software, websites, and services. Rumors abound about several of the leading service offerings: high prices, oversold benefits, promised but undelivered functionality, and difficulties with technical support. To some extent, these are libraries’ standard gripes about all vendor software offerings. The truly unfortunate problem, however, is that the difficulty of migrating existing items from any system or service to another – when migration is even possible, that is – locks libraries into an initial decision that in hindsight may have been a poor one.

Faculty needs

Repository software serves observed and stated faculty needs surrounding content creation and dissemination *hardly at all*. “Roach motel” repositories, in which materials fixed in their final form are the only acceptable content, hold no value for many faculty, which inevitably means such repositories have no access to most faculty-created content. “There is much more experimentation, however, with regard to means of in-

progress communication, where single means of publication and communication are not fixed so deeply in values and tradition as they are for final, archival publication” (Harley et al. 2007). Although much of that content may well be ephemeral, not worth saving, or poorly-suited to preservation, ignoring it altogether ignores the social-engineering advantages to inserting the institutional repository (broadly conceived) into the wider web of faculty content-creation practices. Many faculty could find a use for document versioning, and they frequently express the need to share raw materials such as drafts and datasets with a select few while research is in progress. If the repository does not allow faculty to do these things, *they will not use the repository*, not even for the final product. A two-layered repository approach – one layer for collaboration, one for preservation – is being tested in Australia (Treloar and Groenewegen 2007) and bears watching.

“Research computing” and “cyberinfrastructure” initiatives arising on the campuses of large research institutions are examining efficient and effective Internet-facilitated collaboration and research production (see Gold 2007 for summary). These initiatives rarely consider long-term data curation part of their mandate, however, and when they do, initiative managers are not as educated about the issues as might be wished. Existing institutional-repository software and services cannot seriously be considered the solution, unfortunately (Tonge & Morgan 2007). First, they are too rigid to serve standard faculty needs, such as file versioning and fine-grained access control. Furthermore, they are insufficiently scalable and flexible to deal with the immense quantities of data – even weeded and winnowed data – that research in many disciplines produces. In fact, the widespread failure of institutional repositories may pose a serious danger to library investment in cyberinfrastructure initiatives, either because once-burned administrators are twice shy about further investment in faculty-focused digital initiatives, or because they honestly believe that supporting a repository is adequate response to the need.

The notion that faculty members will actually push buttons and type metadata in order to deposit materials into IRs is an article of faith among repository-software developers. In practice, however, most deposits are third-party mediated, many by librarians, some by faculty members’ support staff or IT personnel. Neither DSpace nor EPrints handles the licensing of mediated deposits appropriately; both insist that the person pushing the buttons is the same person who has authority to grant the license. In

practice, this means either that mediated deposits are not appropriately licensed, or that the DSpace administrator must invent a clunky, often paper-based workflow to secure a license from the appropriate party. Similarly, no workflow exists in either tool to license many deposits at once, making batch-import licensing a tremendous hassle. Ideally, licensing would not occur on a per-item basis at all, but would be governed by a Terms of Service agreement that could be assented to once, never again to interrupt the deposit process. None of the three leading open-source repository packages currently allows this approach.

Moreover, the development of repository software has not been sensitive to faculty's need to document successes. Neither DSpace nor Fedora builds in the sort of sophisticated access logging and tracking that several disciplinary repositories such as the Social Science Research Network have, leaving many institutional repositories at a dreadful disadvantage relative to these more sophisticated repositories in the struggle to attract and keep faculty loyalty. A useful statistics system must count accesses per author and per collection as well as per item and per file, and must provide information about these accesses over defined periods of time. It should track referring links and reader locations as well, so that authors can watch their work spread and participate in online conversations about it. It should also filter out otiose accesses such as those from search-engine crawlers. EPrints has done much better at fulfilling this need; their developers have released a package called IRStats (<http://trac.eprints.org/projects/irstats/wiki>) that works with both EPrints and DSpace and offers most necessary functionality.

The line in the sand drawn between research output and learning objects also acts to the detriment of faculty investment in institutional repositories. Learning objects are immediate-use, highly salient materials in the mind of a teaching faculty member because of the inexorable march of the academic calendar. When the repository refuses them, they go somewhere else – and so does the attention of the faculty member, who promptly files the repository under “one more unnecessary system.” Similarly, faculty are crying out for better records-management and career-documentation systems. Crucially, they want all these systems to be the same system – or at least, to be available from the same portal and to present roughly similar user interfaces. The institutional repository refuses to fill these needs, or to fit into a larger system designed to fill them.

Repositories also need to allow authors to embargo their content temporarily and permanently. When repositories say “deposit is irrevocable, changes are impossible, and access is invariably open,” faculty turn away. Pre-publication collaboration, publisher embargoes, patent holds, and similar issues demand granular access permissions; the collection-level permissions structure in DSpace is irreparably broken for most faculty uses. Permissions need to be set per item during the ingest process, and furthermore must be easily changed, either by faculty fiat or on a schedule determined at ingest.

The fundamental subtext to the question “Can I limit access?” is “Do I control my work?” The small but growing movement toward faculty senates ratifying copyright-retention author’s addenda may in fact express a desire for control over research output. That very desire makes embargoes a somewhat paradoxical recruitment tool for institutional repositories. Although faculty typically embargo their work out of unfounded fears, over time they will deposit more and embargo less simply because the embargo enabled them to overcome their fear long enough to dip their toes in the water. The trajectory of embargoed electronic theses and dissertations at pioneer ETD collector Virginia Polytechnic University followed precisely this pattern (Gail MacMillan, personal communication). EPrints currently allows embargos, and Fedora permits all sorts of access controls to be built atop it. DSpace, however, must be hacked to allow item-level (rather than collection-level) embargos as well as automatic lifting of temporary embargos; for an out-of-the-box installation, only the DSpace administrator can embargo individual items, and that only manually.

Development practice

Open-source “bazaar-style” development practices have been honored more in the breach than in the observance in repository software development. Despite lip service to “the community,” each of the three nominally open-source repository software projects is tightly controlled by an inner circle of developers. In practice, this means that development is slow because of the paucity of developers, and unresponsive because of a dearth of input from the userbase. DSpace in particular has been harmed by its close association with MIT and concomitant overcommitment to solving MIT’s institution-specific issues. As recently as July 2007, a design decision regarding the permissible number of steps in the submission process was justified on the basis of the default

DSpace visual design created by and for MIT, with no acknowledgment that other institutions may have fewer design constraints (<http://www.mail-archive.com/dspace-tech@lists.sourceforge.net/msg01661.html>).

Worse yet, the architectures of these software packages are deeply innovation-unfriendly. DSpace and EPrints are heavily overengineered and written in Java and Perl respectively, rather than one of the simpler Web-scripting languages such as PHP or Ruby. Even slight modifications are out of reach of most members of “the community” due to lack of specialized expertise and steep learning curves. The problem is assuredly *not* that the community is unwilling to pitch in. The impressive number of translations for the DSpace user interface, the creation of which entails the alteration of one simple text file, hints at considerable interest in tasks whose technical demands are minimal. For anything else, the learning curve is just too steep. Even “skinning” the DSpace application – creating a new visual design for it – currently requires either tedious hackwork in JSP and Java code files (for the JSP user interface) or significant XSLT skills (for the newer, soon-to-be-standard Manakin interface).

Neither DSpace nor EPrints was originally designed with a plugin API to allow third parties to modify and expand the package’s functionality. In practice, that means that even though it is *possible* and *legal* to modify the code for local needs, or to use modifications developed by a third party, doing so is deeply *unwise* from a sysadmin’s perspective: code changes require time-consuming re-evaluation, patching, and possibly modification at every new release of the software. It is even possible that a new release may change the codebase so deeply that an older modification is impossible to use without a complete rewrite. Since DSpace and EPrints are so young, and their underlying architectures are not yet fixed, new releases occur yearly or even more often, making the redevelopment burden a genuine concern for would-be innovators.

Innovation has therefore languished. Most repository implementors create local code modifications; almost none share them. Several useful and interesting projects have withered because they required so much code-change that few dared adopt them. Just in DSpace circles, the University of Rochester’s Researcher Pages add-on (Foster and Gibbons 2005), the Tapir project for electronic theses and dissertations (Jones 2004), and Tim Donohue’s automated format-conversion tool (Donohue 2007) have not been widely adopted despite their obvious utility, and all but the last-named are no longer compatible with the current DSpace release. Both DSpace and EPrints are working on

APIs, and EPrints has gone a step further toward web services and import/export plugins, but while this work drags on, innovation languishes further.

Fedora, on the other hand, is a repository *framework* rather than a full-fledged repository; just to *begin* working with it requires substantial up-front development effort to build a data model and develop and bolt on a user interface. Since no two developers building Fedora repositories come up with the same interface ideas or underlying data models, Fedora innovations are difficult to share and generalize, leading (somewhat ironically) to the same innovation paralysis gripping DSpace and EPrints. Generic user interfaces such as the Fez project (http://dev-repo.library.uq.edu.au/wiki/index.php/Main_Page) are slowly being developed for Fedora, but they are even less mature than DSpace and EPrints, and their future is not assured. The community is considering a set of defined “asset actions” to make interoperation easier (Chavez et al. 2007), but agreeing upon and implementing them will take years.

In general, EPrints has been considerably more responsive to faculty needs than either DSpace or Fedora. Its statistics reporting is excellent, its embargo-handling exemplary, and it regularly adds small but useful bonus features such as the email-the-author button on embargoed content. Its ingest interface configures itself to match the type of content deposited (whereas DSpace presents otiose fields to depositors unless a repository manager designs away the problem via a complex XML file), and its recently-developed AJAX-based interface is responsive and pleasant to work with. EPrints is not perfect, of course – its inattention to mediated deposit is a problem – but for a service intended to appeal to faculty, it is currently the best available choice.

User feedback and usability

Troubling aspects of the open-source development experience are alive and well among IR software projects: condescension verging on contempt toward the less-technically-inclined and a vitiated sense of design and usability.

The classic open-source project is developed to “scratch the developer’s own itch,” not to please end-users. Usability, documentation, polish, and end-user support are afterthoughts, if they are on the agenda at all, and open-source developers are legendary for verbal abuse of third parties who run into problems with their software or

offer suggestions (rather than actual code) for improving it. Since it is a developer's playground, useless without significant up-front investment in additional software development, Fedora bypasses the worst of these problems. EPrints intelligently created an in-house paid vendor-support option (<http://www.eprints.org/services/>); DSpace has only an unvetted, out-of-date list of companies on its website (http://www.dspace.org/index.php?option=com_content&task=blogcategory&id=50&Itemid=103), and Fedora claims a single support vendor whose website only mentions a software suite the vendor has built atop Fedora, rather than more broad-based technical support options.

DSpace in particular has had problems managing software support, problems exacerbated by its market position as "out-of-the-box solution." Despite DSpace's substantial userbase, the DSpace technical mailing list is dominated by only a dozen or so voices, and many questions appear to go unanswered, though fewer than in the past. Answers, when they arrive, may take the form of abuse. When the University of Calgary published about its rejection of DSpace for an e-thesis collection (Atkinson 2006), the reaction of DSpace developer Richard Jones (2006) was swift and contemptuous: "DSpace is an Open Source product, where words like 'cannot' should not be used unless you really have looked into it. The underlying search engine can do all of the things required for Calgary, and all it requires is the alteration of the UI to support it." Whatever DSpace's actual capabilities, expressing them in this fashion does not inspire confidence in or fondness for the software.

Some of the usability problems with repository software interfaces will be familiar to anyone who studies human-computer interaction. The out-of-the-box DSpace deposit interface is ugly, complex, overly "clicky," and relentlessly unhelpful to the novice depositor. For example, the interface tells a depositor that the publication year is required in the metadata for a previously-published item – but not until *after* the depositor has omitted the year the first time; nothing on the initial metadata page indicates which metadata fields are required! Once an item has been archived, its depositor cannot change it or its metadata, nor can it be withdrawn or deleted by anyone short of the DSpace sysadmin. This is, to say the least, an overly optimistic view of human capacity for perfection on the first try, a view that forces DSpace sysadmins to waste considerable time cleaning up other people's errors.

Out-of-the-box browse and search interfaces in DSpace incorrectly choose “OR” rather than “AND” as the default search connector, use buttons for scoped browsing when links are more appropriate, display title lists in an ugly table with date rather than title in the first column, and do not help users find related items by making author names and subject keywords clickable on item-display pages. None of these problems is insoluble, but all have persisted through several releases of the software; usability is not a developer itch.

Library IT management

To some extent, these development and interface problems stem from a misunderstanding on the part of developers about typical models of IT management in libraries. Vanishingly few librarians are trained, effective software developers for several reasons, among them librarianship’s notoriously low salaries as well as many librarians’ distrust of computer technology. Most library server applications, then, are not run by librarians, but by dedicated IT staff reputed (often with justification) to be unresponsive to librarian requests. Even librarians competent to make system changes (and librarianship has more sysadmins than developers) may not be *allowed* to because of lines in the sand drawn long ago between librarians and IT staff. Options and functionality buried in configuration files on the server without any sort of web-based interface are wholly unavailable to many librarian repository managers; unfortunately, this is precisely where most options and functionality reside.

The visual design of DSpace-based repositories is a case in point; most adhere very closely to the default design created originally for MIT, right down to the wording in the navigation sidebar. Since the slightest alteration to a DSpace design requires placing files on the server and then restarting the underlying Tomcat software, a librarian without permission to perform these tasks or authority to ask that they be performed on her behalf cannot meaningfully alter her DSpace instance’s appearance or behavior.

Addressing this issue will require considerable developer ingenuity; configuration files are many, of diverse types (some XML, some plain-text, some command-line options, some governed by other system software), and dangerous to meddle with. Possibly exposing at least some configuration options via a bare-bones web interface like that produced by Firefox’s “about:config” command would be a start.

The pathetically inadequate state of repository software undoubtedly contributes to the poor image of librarian repository managers among open-access advocates and faculty at large. When available services are so tightly bound to substantially useless software packages, poor perceptions of the software inevitably leak onto those managing and advocating it.

What must happen

“I can put all that in? That’s great! Why haven’t I heard of you before?” (Faculty member, public policy)

Successes to date – and nothing in this article should suggest that there have been none – point out a number of roads forward for institutional repositories. None of these roads is easy; all require resources, commitment, and will.

Support for institutional repositories begins at home, in libraries, library organizations, and library schools. SPARC and ACRL need to step up efforts to educate present and future academic librarians – regardless of professional specialty – about open access and institutional repositories; the current ignorance is deplorable and damaging. Library-internally, administrators need to make library staff aware of the issues and engage them in repositories as part of the solution. Library schools need to add open access and institutional repositories to the curriculum, and their faculty are honor-bound to self-archive. Likewise, no academic library should sponsor an institutional-repository without first instituting a deposit mandate for professional articles and presentations from staff librarians.

Institutional repository projects should not be launched without at least one significant constituency. It is perfectly reasonable for electronic theses and dissertations or other library-internal projects to spur development of an institutional repository; it is likewise only sensible to start one in response to faculty request, or with the explicit intent of actively mediating deposit on faculty behalf. Any library that cannot or will not start by developing a repository constituency with real commitment to depositing material on an ongoing basis, however, should not start a repository at all.

Institutional repositories should look beyond open access to peer-reviewed literature. This is an otiose commandment, as most repositories already do this to justify their continued existence in the absence of voluntary or mandated self-archiving. However, they do not do it systematically, since there has been no consideration of collection-development

policy. The exact boundaries of desired content will vary among institutions, but likely candidates include digital multimedia, gray literature, learning objects, website capture, digitized special-collections materials, and perhaps even records management (though current repository software is extremely poorly-suited to that). Libraries whose support for repositories rests purely on hopes of collecting peer-reviewed literature would be well-advised not to bother with them.

Institutional repositories should be active, not passive, collectors of content. This is unquestionably a bitter pill, raising serious questions about funding, staffing, and better workflow automation. Still, library administrators unwilling to put staff behind mediated deposit should ask themselves what their repository is intended to accomplish, and whether any other means will realistically meet goals.

Institutional repositories should seek forgiveness rather than permission from faculty. Seeking permissions from faculty for materials already available on the open Web or covered by publisher policy is a huge drain on staff time and a serious stumbling-block for active as well as automated content development. Repositories need to accept the risk that some faculty and perhaps even some publishers will be annoyed by deposits that do not ask their permission first, develop policy around it, and not let those few faculty prevent content collection from the many who do not mind deposit on their behalf but will not sign licenses or handle deposits themselves. Systems that seek one-time consent to a blanket Terms of Service agreement rather than insisting upon separate licenses for every item deposited would be a great step forward.

Institutional repositories should be able to digitize analog content. Our faculty straddle the analog-digital divide, and will continue to do so for some decades. A sensible service does not limit itself to born-digital materials, nor does it expect faculty to manage digitization on their own. E-reserves departments are natural allies in this endeavor.

Institutional repositories should agitate for deposit and harvest APIs in disciplinary repositories, as well as license to harvest them. Faculty preference for disciplinary repositories is natural. Repositories should work with that preference by mirroring material deposited by their institutions' faculty into disciplinary repositories, and insisting that those repositories be responsible academic citizens by allowing such mirroring. Moreover, as funder mandates become more cumbersome, a campus service automating the deposit process into disciplinary repositories (and incidentally snagging

a copy for the institutional repository) should make friends quickly among beleaguered faculty.

Institutional repository managers should involve themselves in discussions of campus cyberinfrastructure, and prepare themselves to help faculty with data management throughout the research cycle. This will mean considerable self-education on data curation, a specialized skill many repository managers do not yet dominate. It will also require considerable education for library administrators, since “[l]ibrarians in particular have not traditionally been involved in the production of scientific information prior to the publication of results.” (Gold 2007) Managers should also lean hard on repository software developers for “integrated information architectures, which link institutional repository and data centre software platforms.” (JISC p. 50) JISC’s RepoMMan project (<http://www.hull.ac.uk/esig/repomman/>) is worth study and imitation, as are efforts at Monash University (Treloar et al. 2007) and Leeds University Library (Stanley 2007). Fundamentally, libraries have no right to demand the final products of a process they have studiously ignored; moreover, the more access they have to the process, the more access they have to the final products thereof.

Institutional repository managers should assert themselves with software developers, and contribute to development however possible. Young though repository software packages are, they need not be so *bad*. Starting at once, we must insist on usable APIs, librarian-accessible configuration, and respectful developer attention to our experiences running these services. Chavez et al. (2007) suggest a wise list of fruitful services to offer, but the software is years behind accomplishing them.

Institutional repositories should be sensitive to faculty prestige needs. At a minimum, this means reliable access statistics as previously discussed. It also means allowing a department’s repository collections to align themselves with the design of the department’s regular website, and playing nicely with institutional-bibliography efforts, even to the point of storing citations without content.

Institutional repositories should both accept content from and disseminate content to standard and specialized campus IT systems. That institutional repository managers cannot build bridges with other data providers and managers on campus seriously threatens repositories’ viability. It is ludicrous that courseware such as Blackboard and Moodle, Web-management tools such as Xythos, and publishing tools such as Open Journal Systems cannot feed into repositories with a few clicks. It is equally ludicrous that

repositories cannot offer transclusion of by-author listings onto faculty websites with one line of Javascript. The SWORD initiative (<http://www.ukoln.ac.uk/repositories/digirep/index/SWORD>) should be watched, as it will provide a bridge for repositories to accept content from other sources. Likewise, Imperial College London's effort to embed the repository in existing campus practice and technology (Afshari and Jones 2007) is exemplary.

We cannot keep looking the other way, pining after mandates we cannot realistically achieve, hoping against hope that the great faculty behemoth will awaken from slumber. England and Australia have accepted that voluntary faculty-initiated and faculty-performed self-archiving is not a viable model for institutional-repository population, and they are beginning to move on. We must do the same. An institutional repository is a useless excrescence unless it is part of a systematic, broad-based, well-supported data-stewardship, scholarly-communication, or digital-preservation program.

References

- Afshari, F., and Jones, R. (2007). Developing an integrated institutional repository at Imperial College London. *Program: electronic library and information systems* 41(4), 338-52. Retrieved October 10, 2007 from <http://spiral.imperial.ac.uk/handle/10044/1/493>
- Atkinson, L. (2006). *The rejection of D-Space [sic]: Selecting theses database software at the University of Calgary archives*. Retrieved August 27, 2007 from <http://eprints.rclis.org/archive/00007938/>
- Bailey, Jr., C. (2007). *ARL publishes Scholarly Communications Education Initiatives SPEC Kit*. Retrieved November 6, 2007 from Charles W. Bailey Jr. <http://digital-scholarship.org/digitalkoans/2007/11/03/arl-publishes-scholarly-communication-education-initiatives-spec-kit/>
- Brody, T., Carr, C., Gingras, Y., Chawki, H. Harnad, S., and Swan, A. (2000, August). Incentivizing the open access research web. *CTWatch Quarterly*. Retrieved August 20, 2007, from <http://www.ctwatch.org/quarterly/print.php?p=86>
- Brown, L., Griffiths, R., and Rascoff, A. (2007). *University publishing in a digital age*. Retrieved October 11, 2007 from <http://www.ithaka.org/strategic-services/Ithaka%20University%20Publishing%20Report.pdf>
- Carr, L. (2007). Size isn't everything: Sustainable repositories as evidenced by sustainable deposit profiles. *D-Lib Magazine* 13(7/8). Retrieved August 20, 2007, from <http://www.dlib.org/dlib/july07/carr/07carr.html>

- Chavez, R., Crane, C., Sauer, A., Babeu, A., Packel, A., and Weaver, G. (2007). Services make the repository. *Journal of Digital Information* 8(2). Retrieved September 28, 2007 from <http://journals.tdl.org/jodi/article/view/195/179>
- Donohue, T. (2007). *Format conversion in DSpace*. Retrieved August 21, 2007, from <http://hdl.handle.net/2142/209>
- Foster, N.F, and Gibbons, S. (2005). Understanding faculty to improve content recruitment for institutional repositories. *D-Lib Magazine* 11(1). Retrieved August 20, 2007, from <http://www.dlib.org/dlib/january05/foster/01foster.html>
- Gold, A. (2007). Cyberinfrastructure, data, and libraries, part 1: A cyberinfrastructure primer for librarians. *D-Lib Magazine* 13(9/10). Retrieved September 18, 2007 from <http://www.dlib.org/dlib/september07/gold/09gold-pt1.html>
- Harley, D., Earl-Novell, S., Arter, J., Lawrence, S., and King, C. (2007). The influence of academic values on scholarly publication and communication. *Journal of Electronic Publishing* 10(2). Retrieved October 14, 2007 from <http://hdl.handle.net/2027/spo.3336451.0010.204>
- Harnad, S. (2006). Two happy accidents demonstrate power of “eprint request” button. Retrieved August 20, 2007 from <http://openaccess.eprints.org/index.php?/archives/167-Two-Happy-Accidents-Demonstrate-Power-of-Eprint-Request-Button.html>
- Hendler, J. (2007). Reinventing academic publishing – part 1. *IEEE Intelligent Systems* 22(5), pp. 2-3
- Hitchcock, S. (2007). *The effect of open access and downloads ('hits') on citation impact: a bibliography of studies*. Retrieved August 28, 2007, from <http://opcit.eprints.org/oacitation-biblio.html>
- Jones, R. (2006). *Calgary rejects DSpace for e-theses archives*. Retrieved August 27, 2007, from <http://chronicles-of-richard.blogspot.com/2006/12/calgary-rejects-dspace-for-e-theses.html>
- Jones, R. (2004). *The Tapir: Adding e-theses functionality to DSpace*. Retrieved August 21, 2007, from <http://hdl.handle.net/1842/632>
- Lyon, L. Dealing with data: Roles, rights, responsibilities and relationships. Retrieved October 4, 2007 from http://www.jisc.ac.uk/media/documents/programmes/digitalrepositories/dealing_with_data_report-final.pdf
- McDowell, C. (2007). Evaluating institutional repository deployment in American academe since early 2005: Repositories by the numbers, part 2. *D-Lib Magazine* 13(9/10). Retrieved September 18, 2007 from <http://www.dlib.org/dlib/september07/mcdowell/09mcdowell.html>
- Medeiros, N. (2007). The catalog's last stand. *OCLC Systems and Services* 23(3), 235-7.
- Poynder, R. (2006a). The OA interviews: Stevan Harnad. Retrieved October 4, 2007 from <http://poynder.blogspot.com/2007/07/oa-interviews-stevan-harnad.html>
- Poynder, R. (2006b). Clear blue water. Retrieved October 4, 2007 from <http://poynder.blogspot.com/2006/03/institutional-repositories-and-little.html>

- Pryor, G. (2007). Attitudes and aspirations in a diverse world: The Project StORe perspective on scientific repositories. *International Journal of Digital Curation* 1(2). Retrieved October 4, 2007 from <http://www.ijdc.net/ijdc/article/view/32/35>
- Sale, A. (2007). The patchwork mandate. *D-Lib Magazine* 13(1/2). Retrieved August 20, 2007, from <http://www.dlib.org/dlib/january07/sale/01sale.html>
- Stanley, T. (2007). Developing a virtual research environment in a portal framework: The EVIE project. *Ariadne* 51. Retrieved October 14, 2007 from <http://www.ariadne.ac.uk/issue51/stanley/>
- Swan, A. (2006). The culture of open access: researchers' views and responses. In N. Jacobs (Ed.), *Open access: key strategic, technical, and economic aspects* (pp. 65-72). Oxford: Chandos Publishing.
- Tonge, A., and Morgan, P. (2007). *Project SPECTRA: Submission, preservation, and exposure of chemistry teaching and research data*. Retrieved October 14, 2007 from <http://www.lib.cam.ac.uk/spectra/FinalReport.html>
- Treloar, A., and Groenewegen, D. (2007). ARROW, DART, and ARCHER: A quiver full of research repository and related projects. *Ariadne* 51. Retrieved October 14, 2007 from <http://www.ariadne.ac.uk/issue51/treloar-groenewegen/>
- Treloar, A., Groenewegen, D., and Harboe-Ree, C. (2007). The data curation continuum: Managing data objects in institutional repositories. *D-Lib Magazine* 13(9/10). Retrieved September 18, 2007 from <http://www.dlib.org/dlib/september07/treloar/09treloar.html>
- University of California Office of Scholarly Communication. (2007). Faculty attitudes and behaviors regarding scholarly communication: Survey findings from the University of California. Retrieved October 14, 2007 from <http://osc.universityofcalifornia.edu/responses/materials/OSC-survey-summaries-20070828.pdf>
- Vincent, C. (2007, July 16). The purloined bibliography. *Chronicle of Higher Education*.
- Willinsky, J. (2005). *The access principle*. Boston: MIT Press.
- Wren, J. D. (2005). Open access and openly accessible: a study of scientific publications shared via the internet. *BMJ*, 330, 1128.