SPARC
Institutional Repository
Checklist & Resource Guide

The Scholarly Publishing & Academic Resources Coalition
21 Dupont Circle • Washington, DC 20036
www.arl.org/sparc

Prepared by Raym Crow, SPARC Senior Consultant

Acknowledgements
SPARC and the author wish to thank the many people whose comments have shaped this document. We would especially like to thank Chris Gutteridge of the University of Southampton and EPrints; Ed Sponsler and Kim Douglas of the Caltech Libraries; and MacKenzie Smith of the MIT Libraries and DSpace. The benefit of their experience and insight has improved this *Guide* immeasurably. Any errors or omissions that remain, however, are the sole responsibility of SPARC and the author.

This document will be revised and updated on an ongoing basis as new information warrants. To comment on this document, please contact Raym Crow at raym@arl.org.

SPARC
Institutional Repository
Checklist & Resource Guide

Release 1.0
November 2002

TABLE OF CONTENTS

INTRODUCTION

The Institutional Repository

Institutional repositories—used in this *Checklist & Guide* to indicate digital collections capturing and preserving the intellectual output of a single university or a multiple institution community of colleges and universities—provide a compelling response to two strategic imperatives of most academic institutions. Such repositories:

- Provide a critical catalyst and component in reforming the system of scholarly communication by expanding access to research, reasserting control over scholarship by the academy, and bringing heightened relevance to the institutions and libraries that support them; and
- Have the potential to serve as tangible indicators of an institution's quality and to demonstrate the scientific, societal, and economic relevance of its research activities, thus increasing the institution's visibility, status, and public value.

Institutional repositories contribute as a logical extension of a university's core mission and as a channel through which to increase institutional visibility. However, they can achieve far greater results in synergy with a network of interoperable open access repositories. Further, they build on a growing grassroots faculty practice of self-posting research online. While institutional repositories necessitate that libraries—as their logical administrative proponents—broaden both faculty and administration perspectives on a wide range of issues, they can be implemented without radically altering the *status quo*. Moreover, they can be introduced by reallocating existing resources, usually without extensive technical development.

Institutional repositories provide practical opportunities to increase faculty and administrator awareness of author rights and larger intellectual property issues, and provide faculty-authors and administrators with specific paths of action to contribute positively to—and benefit from—alternative scholarly publishing channels. In sum, institutional repositories offer a strategic response to systemic problems in the existing scholarly journal system—and the response can be applied immediately, reaping both short-term and ongoing benefits for universities and their faculty and advancing the positive transformation of scholarly communication over the long term.

Document Purpose

The SPARC *Institutional Repository Checklist & Resource Guide* provides an overview of the major issues that institutions and consortia need to address in implementing an institutional repository. These issues include:

- Organizational, administrative, and cultural issues;
- Content policies and accession and retention policies;
- Faculty outreach and participation; and
- Technical options and infrastructure issues.

This *Checklist & Guide* complements SPARC's *Position Paper*, which addresses the strategic implications of institutional repositories.[1]

This document does not provide step-by-step instructions on establishing an institutional repository. Given the vast range of administrative, political, cultural, financial, and technical variables at the institutions and consortia that will be interested in implementing

[1] Crow (2002).

repositories, a detailed manual guiding each step and accounting for all possible variables would be virtually impossible to produce.[2] Rather, the *Checklist & Guide* provides a contextual introduction to each of the issues that one might consider in a particular institution's context, and directs readers to resources that provide additional detail. Combining these with the SPARC position paper, our hope is to provide an effective resource to help guide the planning and implementation of your institutional repository project.

This *Checklist & Guide* relies, whenever possible, on the experiences of those who have actually implemented institutional repositories. We point to those sources throughout this document, particularly in the "Resources & Further Reading" text at the end of each section. Presenting those valuable resources in a topical fashion will make it easier for readers to find information on a particular topic and to compare and benefit from the experiences of various groups. In referring readers to additional resources, we have striven to identify what we believe are the best and most relevant, not to provide a comprehensive list of every resource that may be available.

Intended Audience

The intended audience for the *Checklist & Guide* includes librarians, faculty, administrators, information technology and support staff, and others interested in the practical implications of an institutional repository. Our topical overviews reflect the assumption that readers have a general awareness of the current issues surrounding scholarly communications but have little or no in-depth exposure to the specific topics being discussed here. (Those already familiar with a particular topic may still want to refer to the additional resources suggested.)

We have also tried to avoid projecting the fallacious impression that one effective response exists for every repository implementation issue that might arise. There are few definitive solutions pertinent in all situations. Given the incipient stages of institutional repository adoption, many of these issues are just now being addressed for the first time in various institutional settings, sometimes with initial success, sometimes requiring multiple attempts. As the experience of those implementations adds to our understanding, SPARC will update this document and continue to publicize new developments of community-wide interest.

Much of the information presented here should prove of interest to both individual institutions and to institutions participating in a collaborative consortium implementation. However, issues that pertain uniquely to a consortial implementation lie beyond the scope of this document. SPARC hopes early consortia adopters will remedy this exclusion by providing complementary documents to supplement this *Checklist & Guide*.

SECURING ADMINISTRATION SUPPORT & FACULTY PARTICIPATION

Institutional repositories offer considerable benefits to the institutions that sponsor them and to the faculty, researchers, students, librarians, and others that participate in them. At the same time, institutional repositories might encounter resistance from administrators, faculty, and others who either fail to understand the benefits that such repositories can deliver or who fear that institutional repositories will have a deleterious impact on the

[2] Such an approach might well prove possible for proponents of particular technical solutions. For example, EPrints and DSpace—to cite two current examples of institutional repository systems—are preparing documents to guide implementation of their particular solution. Such guides will complement the information presented in this *Checklist & Guide*.

current journal publishing system, a critical driver of academic advancement. As many faculty and administrators are heavily invested in these systems, and consider their perpetuation essential, the clarity with which proponents communicate the benefits of institutional repositories to these key participants will prove critical. Equally, understanding and systematically addressing the objections raised to repositories will prove crucial to faculty participation and to the ultimate success of each repository implementation.

Securing Administration Support

The perceptions and attitudes of university administrators are critical to gaining the support necessary to validate a repository's standing within an institution. Even where a repository is implemented and managed entirely as a library initiative, the nature and extent of the efforts required to gain faculty awareness and participation in the repository presuppose the buy-in of an institution's administration and its willingness to reallocate resources and/or provide additional funding.

The rationale for universities and colleges implementing institutional repositories rests on two interrelated propositions: one that supports a broad, pan-institutional effort and another that offers direct and immediate benefits to each institution that implements a repository.

New Scholarly Publishing Paradigm

While institutional repositories centralize, preserve, and make accessible an institution's intellectual capital, at the same time they will—ideally—form part of a global system of distributed, interoperable repositories that provides the foundation for a new disaggregated model of scholarly publishing.

As producers of primary research, it is only to be expected that academic institutions would take an interest in capturing, disseminating, and preserving the intellectual output of their faculty, students, and staff. Traditionally, scholarly publishers and academic libraries served complementary roles in facilitating scholarly communication and preserving an institution's intellectual legacy. Over the past several decades, however, the rate of change to the economic, market, and technological infrastructures that sustained this symbiotic publisher-library relationship has begun to accelerate.

These changes—including evolving digital publishing technologies and expanding networking; significant increases in the volume of scientific research; decreasing satisfaction with traditional journal price and market models; and uncertainty over responsibility for long-term digital preservation of scholarly materials—have evolved and combined to create new expectations in the academic community for the production, distribution, and interchange of scholarly communications. In such an environment, institutional repositories might well act to preserve an institution's intellectual work product while contributing to a fundamental, long-term change in the structure of scholarly communication.

Institutional Visibility and Prestige

The responsibility for communicating an institution's strengths, and positioning the institution with the broader context of its markets or communities and funding sources (whether public or private) falls to the institution itself. Obviously, however, merely stating that an institution is committed to academic excellence and scientific progress does not prove the assertion. An institutional repository and supporting metrics provide university administrators with demonstrable evidence of the institution's quality. Institutional repositories help university and college administrators—including Development and Marketing officers—reinforce an institution's brand position and prestige.

Institutional repositories, by capturing, preserving, and disseminating an institution's collective intellectual capital, serve as meaningful indicators of academic quality. Currently, much of each institution's intellectual output is diffused through thousands of scholarly journals. While faculty publication in these journals reflects positively on the host university, an institutional repository concentrates the intellectual product created by a university's researchers, making a clearer demonstration of its scientific, educational, social, and economic value. Thus, institutional repositories complement existing metrics for gauging institutional productivity and prestige. Where this increased visibility reflects a high quality of scholarship, this demonstration of value can translate into tangible benefits, including the funding—from both public and private sources—that derives in part from an institution's status and reputation.

While there is some evidence that administrators and research managers agree that the institution should play an important role in distributing its research output,[3] there are also indications that administrators harbor real concerns about some of the issues raised by institutional repositories. These concerns include:

- The potential impact of institutional repositories on the existing journal publishing system;
- The costs of a parallel system of scholarly communication and the long-term investment required; and
- Intellectual property policy issues.

A marketing communication and education program seeking to persuade an institution's administrators of the value of an institutional repository must address and overcome these potential objections. We outline potential responses—and point to additional resources—below.

Impact of Institutional Repositories on the Existing Publishing System

Many university administrators share faculty-author attitudes (and ambivalence) towards the traditional system of scholarly journal publishing. While recognizing the negative effects of serials pricing on the library's collection and services, university managers and administrators participate in the formal structure of the academic reward system. This system, based substantially on the system of peer-reviewed academic journals, continues to work well for many participants in the system, both authors and administrators.[4]

The resolution of such critical and complex issues, though germane, need not be a prerequisite to initiating an institutional repository. An administration's attitudes towards academic credentialing, its reliance on the existing journal publishing system as a component of academic advancement decisions, and its openness towards alternative methods will likely vary from institution to institution. It is critical here—as elsewhere—to show that institutional repositories augment, rather than displace, the existing system of scholarly journals in providing important new measures of academic performance and in ensuring greater leverage of a particular institution's intellectual capital.

[3] See, for example, the study undertaken by the ARNO project of university administrators and research managers in the Netherlands (Bentum (2001b)). We are unaware of any similar studies of administrator attitudes for North America or other parts of Europe.

[4] See ALPSP (1999), p. 7 and Bentum (2000).

Cost Issues

Given the difficulty of accurately projecting costs for institutional repositories, especially in terms of digital archival preservation, it is understandable that institutional administrators will be apprehensive about the potential long-term expenses. Even where the administration is aware of, and responsive to, the economic burden the institution incurs from rising serials costs, the expense of maintaining a parallel, supplementary system of scholarly communication will doubtless generate debate.

These concerns can be addressed by:

- Positioning the repository as a long-term investment in changing the structure of scholarly communication;
- Presenting the repository as a potential future cost savings as the marketplace responds to institutional initiatives;
- Adducing the direct benefits—both tangible and intangible—that a successful repository delivers to its host institution; and
- Making the case, as diplomatically as possible, that administrators cannot base their decisions solely on financial considerations if the institution is to retain its high stature and reputation for innovation.

These responses are not mutually exclusive and can be applied in combination. However, the first approach assumes concurrence from university administrators on the larger issue of participating in reforming the current system of scholarly communication and the second requires that administrators adopt a long-term perspective on the issue of cost recovery or return on investment in pure economic terms.

Presenting institutional repositories as a long-term investment that helps change the current scholarly communication model—and weaken publisher monopolies on faculty-generated content—presupposes that an institution's administrators understand and agree with that goal. Such an appeal to institutional altruism will be applied with varying effect, depending on the institution and the administrator. In those instances where key administrators—in addition to the library director—are sympathetic to the need to reform the system of scholarly communication, then it makes sense to position repositories as a means to that end. In such a context, a communication program can increase awareness and stimulate discussion of growing economic dysfunctions facing the academic journal publishing system, the impact of these economic issues on the university itself, and a vision of the possibilities of alternative publishing channels. (The Resources section below points to sources for such information.) However, even with sympathetic administrators, when budgets are tight and resources scarce, the exigencies of current and near-term budgeting will tend to work against arguments for long-term investments, particularly for infrastructure improvements that some may consider abstract or non-essential. In this context, the library, as the logical administrative agent for an institution's repository, would have to consider the reallocation of internal library resources.

While there are potential long-term savings over the current system of periodical subscriptions, there is little prospect for substantial, immediate cost reductions. Nevertheless, institutional repositories can be positioned as an active response to the serials price issue, even if immediate economic benefits are not forthcoming. As with the scholarly communication reform issue just addressed, this argument plays, at best, a supporting role to the direct benefits that repositories can deliver.

As we have discussed, communicating the direct benefits that an institution would enjoy from a repository will typically provide the most effective argument for immediate action

(whether seeking authorization to research a proposed initiative further or approval of an actual implementation plan). A direct and immediate benefit is the contribution an institutional repository makes to institutional prestige and visibility, as described above. In this respect, institutional repositories are comparable to the investment that some institutions have made in strengthening academic departments, or in expanding their university presses, which also reflect on the stature of the institution.

Speaking in terms of the benefits derived from increased institutional visibility should have a more immediate impact on administrator perceptions than the secondary benefits discussed above. These benefits include:

- In the U.S. and some other countries, government funding for institutions that receive such public assistance.
- Fundraising and development efforts for both private and public institutions.
- In some European countries, the impact of research made available through e-print[5] servers and institutional repositories is considered in qualitative evaluations of programs and individual faculty.
- In the U.K., institutional repositories might prove useful in managing submissions for future Research Assessment Exercises by ensuring that a good number of papers are easily available in advance.[6]

Intellectual Property Issues

Many university administrators recognize that their academic constituency comprises both creators as well as users of original intellectual property. Therefore, an advocacy approach must balance these dual concerns. Promoting a balanced approach to intellectual property issues—emphasizing author rights, including the retention of rights for self-archiving and educational purposes (as described below)—should help allay administration concerns. Similarly, gaining the approval and enlisting the support of institutional offices with a vested interest in faculty and institutional copyright issues (for example, the university copyright office and/or sponsored research office) should also help gain administrators' support.

Securing Administration Support

Resources & Further Reading

- SPARC has created a SPARC-IR discussion list, an online forum where participants can ask questions, share best practices and debate relevant issues. To sign up, see: <https://mx2.arl.org/Lists/SPARC-IR/>.
- Maarten van Bentum. "Attitude of Academic Staff and [Research] Managers to Electronic Publishing and the Use of Distributed Document Servers on University Level: A Survey Report." ARNO Report (Work Package 7). November 2001. Available from <http://cf.uba.uva.nl/en/projects/arno/workpackages/arnowp7-survey.rtf>.

  Recognizing the importance of securing the participation of academic authors and research managers (e.g., deans, department heads, and research institute directors) for the success of their cooperative

[5] "E-prints" is used to refer primarily to digital preprints, although it sometimes used to encompass published material (postprints) as well. Additionally, EPrints is also the name for a server software system (see <www.eprints.org>) developed to facilitate the posting and use of e-prints. To minimize confusion, here we will use the term "e-print" to refer to self-archived material, whether preprint or postprint. We will refer to the software system as Eprints or Eprints.org.

[6] See Pinfield, Gardner, and MacColl (2002).

institutional repository initiative, ARNO (Academic Research in the Netherlands Online) surveyed academic authors and research managers to ascertain perceptions about electronic publishing and specifically about the use of institutional servers as a parallel publication channel. ARNO plans to use the understanding gained from its survey as the basis for its public relations and marketing programs to encourage participation in the ARNO repository. While the study's sample is too small to be representative, it does provide a qualitative understanding of some of the concerns facing faculty and research managers across several disciplines.

- Maarten van Bentum, Renze Brandsma, Thomas Place, and Hans Roes (2001) "Reclaiming academic output through university archive servers." New Review of Information Networking (August). Available from <http://cwis.kub.nl/~dbi/ users/roes/articles/arno_art.htm>.
- Malcolm Litchfield. "Presses Must Stress Ideas Not Markets." *The Chronicle Review* (June 28, 2002): B9-B10.
- *Principles for Emerging Systems of Scholarly Publishing*. May 10, 2000. Set of principles to guide the transformation of the scholarly publishing system. Agreed to by academic institutional and library administrators as a result of a meeting held in Tempe, Arizona in 2000, sponsored by the Association of American Universities, the Association of Research Libraries, and the Merrill Advanced Studies Center of the University of Kansas. While providing a consensus on principles, the document does not attempt to articulate practical steps to effect the principles set forth. Available from <http://www.arl.org/scomm/tempe.html>.

SECURING FACULTY PARTICIPATION

At most institutions, faculty participation in the institution's repository will have to be sensitive to the scholars' sense of independence. Thus, it should be voluntary or risk encountering resistance, even from faculty chairs and members who might otherwise prove supportive. Understandably then, the direct benefits of participating in an institutional repository must be articulated clearly, emphatically, sensitively, and frequently to engender faculty enthusiasm and support. Further, as noted above, potential objections to institutional repositories must be understood and adequately addressed to overcome initial faculty resistance to participation.

The greatest obstacle to any change in the fundamental structure of scholarly communication lies in the inertia of the traditional publishing paradigm. Academic authors publish for professional recognition and career advancement, as well as to contribute to scholarship in their discipline. Accommodating these faculty needs and perceptions—and demonstrating the relevance of an institutional repository in achieving them—must be central to content policies, implementation plans, and internal education and advocacy programs.

Primary Author Benefits

While gaining credit for professional advancement is a key motivation for academic publishing, the primary reason is communicating with others about their research and contributing to the advancement of knowledge in their field. The principal author benefit of participating in an institutional repository—enhanced professional visibility—supports this goal well. This visibility and awareness is driven by both broader access and increased use. No library can afford a subscription to every possible journal, rendering much of the research literature inaccessible to many of an institution's researchers. Interoperability protocols and standards, when applied to institutional repositories, create the potential for a global network of cross-searchable research information. By design, networked open access repositories lower access barriers and offer the widest possible dissemination of a scholar's work.

A related author benefit derives from the increased article impact that open access papers experience compared to their offline, fee-based counterparts, whether print or electronic. Research has demonstrated that, with appropriate indexing and search mechanisms in place, open access online articles have appreciably higher citation rates than traditionally published articles.[7] This type of visibility and awareness bodes well for both the individual author and for the author's host institution.

Benefits to Teaching Faculty

Besides the benefits for faculty as authors, institutional repositories also deliver benefits to teaching faculty. By including non-ephemeral faculty-produced teaching material, the repository serves as a resource supporting classroom teaching. These materials might include concept illustrations, visualizations, models, course videos, and the like—much of the material often found on course web sites. This benefit should help extend the appeal of institutional repositories across a broader audience of research and teaching faculty.

Benefits to Faculty

Resources & Further Reading

- Maarten van Bentum. "Author's Attitudes and Perceptions and Strategies for Change with Respect to Electronic Publishing: A Literature Study." ARNO Report (Work Package 7). March 2001. Available from Available from <http://cf.uba.uva.nl/en/projects/arno/workpackages/arnowp7.rtf >.

  A study prepared in September 2000 by the ARNO project, a cooperative undertaking of the libraries of three Dutch universities (Tilburg University, University of Amsterdam, and the University of Twente). A literature survey on faculty attitudes and perceptions of electronic publishing and the posting of research to a university server. Includes a discussion of broad (pan-institutional) strategies to encourage faculty-author participation in such repositories.

- ALPSP. *Authors and Electronic Publishing: The ALPSP research study on authors' and readers' views of electronic research communication*. (The Association of Learned and Professional Society Publishers, 2002).

- Steve Lawrence. 2001. "Online or invisible?" *Nature* 411 (6837): 521. Available at: <http://www.nature.com/nature/debates/e-access/Articles/lawrence.html>.

- Stephen Pinfield, Mike Gardner, and John MacColl. "Setting up an institutional e-print archive." Ariadne 31 (April 11, 2002).
  Article outlines the major issues involved in establishing an institutional repository based on the experiences of the universities of Edinburgh and Nottingham. Both institutions implemented their repositories using EPrints software (release one). Available from <http://www.ariadne.ac.uk /issue31/eprint-archives/intro.html>.

- For examples of how an academic community might use an institutional repository, including teaching support, see the use studies developed by MIT's DSpace project: <http://www.dspace.org/live/implementation/usecase.html>.

- Best Practices example for scholarly publishing. See: <http://www.mat.univie.ac.at/~michor/ceic-best.pdf>.

Addressing Faculty Objections

The positive case made for an institutional repository needs to be balanced by addressing concerns and objections that faculty might raise. Surveys of faculty perceptions and

[7] See Lawrence (2001). In the case of computer science articles that Lawrence studied, online articles were cited 4.5 times more than offline articles.

attitudes, and the experiences of previous repository implementers,[8] have documented that these concerns include:

1) Impediments to publication in a prestigious journal, whether the work is posted to the institutional repository prior to or after formal journal publication;
2) Perceived low status from lack of quality control and peer review;
3) Intellectual property rights, particularly copyright, and information abuse;
4) Undermining of the current system of academic journal publishing; and
5) Added faculty workload to submit content.

As we will see in detail below (see "Impact of Discipline-specific Practices"), the nature and intensity of these objections differs between academic disciplines, and supports the practice of developing content policies tailored to each research community to reduce author skepticism and encourage participation. We will address each of these objections in turn.

Impediment to Publication

Among the most frequently cited concerns of academic authors considering posting research in an institutional repository is the impact that such posting might have on publication in a traditional peer-reviewed journal. As such formal publication remains essential for academic professional advancement, the perception that posting to an institutional repository might preclude journal publication would discourage faculty participation.

In many disciplines, informal methods of pre-publication communication—including preprints, conference presentation, poster sessions, published abstracts—have long been recognized as important and legitimate components of scholarly communication and not considered formal publication. Hence, such dissemination typically did not preclude subsequent formal publication in a peer-reviewed journal. Indeed, one can argue that—from a scholarly communication perspective—posting a research communication to a personal web page or to an institutional repository differs little from presenting the same material at a conference: both allow for comment and revision prior to formal, definitive publication.

There is increasing recognition, at least in the sciences, that scholarly publishing represents such a continuum, and the previous resistance of many journal publishers to prior electronic publication is changing. A number of scientific journal publishers have adopted the position that posting on e-print servers or institutional repositories does not in itself constitute prior publication, but rather provides a legitimate channel of scholarly communication.

Still, publishers in medicine and chemistry (for example, the *New England Journal of Medicine* and the American Chemical Society) continue to maintain stringent prohibitions against prior online posting. Interestingly, however, journals in physics, astronomy, computer science, economics, and demography—which, given the prevalence of e-prints posting among their authors, had to acquiesce in the practice of online posting—seem to have lost none of their prestige or financial strength as a result.[9] The

[8] See ALPSP (2002) and Pinfield, Gardner, and MacColl (2002).

[9] The Institute of Physics author agreement requires authors to cede copyright, but grants a "personal license" "to post and update the Work on non-Publisher servers (including e-print servers) as long as access to such servers is not for commercial use and does not depend on payment for access, subscription, or membership fees." See: <http://www.aip.org/pubservs/authserv.html>.

reason appears to be that authors and readers in those disciplines perceive a qualitative difference between informal and formal publication. Informal publication is considered weaker than the prestige, credibility, and added branded visibility of stronger formal publication.

In practice, publisher policies towards Internet posting of articles prior to or after journal publication vary widely: some journals will consider for publication research previously posted on the Internet and will allow posting of the published work on an author's personal and/or institutional Internet site; others consider the posting of material on the Internet as "prior publication" and forbid author self-archiving and even educational use of the material. Actual policies reflect multiple variations on these themes. On a practical level, the library can work with individual contributors and their publishers to address these issues and seek a mutually satisfactory resolution. To help maintain the distinction between the repository as an informal communication channel and peer-reviewed journals as a formal channel—for the benefit of both faculty and publishers—it would be best to avoid terms such as "submit" and "publish" in referring to faculty contributions, using instead "participate," "deposit," "contribute," or "post."[10]

Concerns About Repository Participation Impeding Publication

Resources & Further Reading

- Eugene Garfield. "Acknowledged Web Posting is Not Prior Publication." *The Scientist* 13 (12): 12 (June 7, 1999). Available from <http://www.the-scientist.library.upenn.edu/yr1999/June/comm_990607.html> (requires free registration)
- Editorial. "What is publication?" *BMJ* Volume 138 (16 January 1999), p.142.
- See Declan Butler. "The writing is on the Web for science journals in print." *Nature*, vol. 397, no. 6716, Jan. 21, 1999, pp. 195-200.
- See the "I Worry About…" FAQs at the EPrints.org site for responses to faculty objections on a number of issues. Available from <http://www.eprints.org/self-faq/>.

Quality Control & Perception

As we have seen above, various versions of research publication serve different purposes in the scholarly communication continuum. In any event, researchers must be confident that research—in whatever published stage, form, or venue—is legitimate and well executed prior to committing time to reading and using it. At the same time, quality control issues—including concern with the commingling of peer-reviewed articles with working papers—present another obstacle to faculty author participation in institutional repositories.

The vast majority of faculty authors, when weighing publishing options prefer to submit articles to journals with formal peer review. Surveys suggest that authors feel strongly about the importance of peer review, editorial selection, quality control, and other components of the traditional journal publishing process. Further, they indicate the reluctance of some faculty to contribute published articles to a repository if they appear alongside non-peer-reviewed material.[11]

Formal peer review is only one process for ensuring quality. As we will discuss, depending on the content policies of a repository's constituent scholarly communities, a repository might contain not only peer-reviewed and non-peer-reviewed material, but

[10] See the eprints.org self-archiving FAQ: <http://www.eprints.org/self-faq/>.
[11] See ALPSP (2002) and Bentum (2001b).

research with intermediate levels of certification. In some disciplines—high energy physics, for example—material from extensive and collaborative research projects often receives considerable internal review prior to, or during, the preprint stage. Some occasional or working paper series explicitly differentiate their contents qualitatively from "published" research, while reflecting relatively strong quality indicators. In such manuscript series, primary certification is inherent in affiliation with a university or research program.[12] Analogously, the multiple academic user communities (for example, departmental faculty or research center fellows) that constitute an institutional repository represent selective and tightly controlled fields of membership. A department's reputation is a function of this selectivity, which in turn correlates to the assumed quality for the department's repository contributions.[13]

Implementing a repository using a user community-oriented content approval structure allows institutional sponsorship and departmental participation to lend legitimacy to the repository's content. Additionally, other repository policies can address these concerns, and combat the perception that institutional repository posting is inherently low status. These policies include:

- Differentiating between preprints and published peer-reviewed research. Including various types of formal and informal scholarly communications is desirable as long as readers are made aware of what they are reading: un-vetted preprint or peer-reviewed article. In the institutional repository context, this translates into the need for utter transparency in letting users know what they are reading. If peer-reviewed and non-peer-reviewed material is included in the same repository, they should be clearly labeled and even maintained in separate areas of the site. This will help users differentiate certified from non-certified content.[14]
- Distinguishing between "self-publishing," which is perceived as vanity publishing, and the "self-archiving" of published, refereed material.

Quality Control and Perception

Resources & Further Reading

- Stevan Harnad. "Five Essential Post Gutenberg Distinctions." Available at: <http://www.ecs.soton.ac.uk/~harnad/Tp/resolution.htm#1.4>.
- Rob Kling, Lisa Spector, and Geoff McKim. "Locally Controlled Scholarly Publishing via the Internet: The Guild Model." *CSI Working Paper* no. WP-02-01 (June 2002).

Intellectual Property Issues & Author Rights

Participation in institutional repositories raises several intellectual property concerns amongst faculty. One relates to the issue addressed above: the concern that posting to an institutional repository will be considered prior publication, hence rendering the author's intellectual property rights to the research essentially worthless.

A second concern is that open access—whether through an institutional repository, personal web site, or other channel—will jeopardize author control of the research and

[12] See, for example: Berkeley Roundtable on the International Economy (BRIE) (<http://brie.berkeley.edu/~briewww/pubs/index.html>); Harvard Business School research manuscript series (<http://www.hbs.edu/dor/papers.index.html>); and University of Western Ontario Population Studies Centre (<http://www.ssc.uwo.ca/sociology/popstudies/dp.html>).

[13] See Kling, Spector, and McKim (2002).

[14] See Stevan Harnad. "Five Essential Post Gutenberg Distinctions." Available from <http://www.ecs.soton.ac.uk/~harnad/Tp/resolution.htm#1.4>.

expose it to plagiarism, misinterpretation (by the media, for example), and other forms of information abuse. Perception of such threats is conditioned, at least in part, by the practices of each discipline. As one might expect, concerns about protecting work-in-progress appear more pronounced in those fields without a tradition of widely sharing such work.[15]

Fear of such information abuse stems, at least in part, from a perception that an institutional repository would exercise inadequate control over the content. It is important, therefore, to ensure that the institution's repository does indeed provide sufficient control, to emphasize the point to faculty, and to engage faculty representatives in designing relevant policies and practices. Faculty authors must have confidence that their research material will not be co-opted by others—who might take the research further, faster, and with greater impact—or even plagiarized outright, thus damaging their prospects for career advancement.[16]

To deserve this confidence, institutional repositories must serve a basic registration function, recording the priority of ideas and intellectual property. The potential and importance of such registration will probably have more impact on some fast cycling, high volume disciplines, although some of these already have their own discipline-specific repositories which also provide a basic registration function.

While printed journals will continue to provide the preeminent venues for registration and certification for the foreseeable future, institutional repositories will allow a greater proportion of researchers to register their work in a recognized forum. However, registration in itself only represents an initial step. Certification, such as peer review, validates the quality of the research and thus confirms the registration of intellectual priority. In addition to basic registration, there will be instances where an academic community (for example, a department, research center, or lab) exercises some level of qualitative content control that serves a certification function analogous to—but rarely as rigorous as—traditional peer review. The validity of the registration is thus, in part, a function of certification quality. However, even without certification mechanisms, the repository can document the date that material is posted and display copyright notices or rights appropriate to the content. Current e-print servers currently provide this level of control and protection, which appears to be sufficient to encourage participation, though again for disciplines comfortable with circulating working papers.

Author Rights

A complementary tack in addressing faculty intellectual property concerns is to promote a fuller understanding of author rights and the benefits of authors retaining rights to their research. While this issue obviously has implications beyond institutional repositories, repository participation provides a logical context for the discussion. In any event, as faculty grow increasingly aware of the value of their intellectual property in other areas, such as distance education, one suspects that they will grow more attentive to their rights in terms of scholarly publishing.[17]

The issue of intellectual property, both in the academy and beyond, is fraught with legal and economic implications. As with other issues, focusing on the direct benefits to

[15] See ALPSP (2002) and Bentum (2001b).

[16] See Kling and McKim (2000).

[17] An increasing number of authors appear to consider it important to retain copyright, even when they continue to sign over full publishing rights to a publisher. This suggests a growing awareness of the copyright issue amongst faculty authors. See ALPSP (2002), p.24.

faculty of retaining certain rights will help focus the issue in a manner relevant to faculty repository participation and help repository implementers avoid broader battles they are disinclined or ill-prepared to fight.

Academic institutions and their faculties should manage copyright in a manner that assures faculty access to use of their published works in research and teaching, while balancing the legitimate business interests of publishers. Neither authors nor publishers need to own copyrights in order to gain the rights necessary to achieve their legitimate goals. For the publisher's part, a rights management arrangement that grants publisher exclusivity for first publication should prove sufficient to enable the publisher to earn a reasonable return on its investment and ensure the journal's viability. For their part, faculty authors should assign the rights to their work in a manner that allows the broadest possible access. At a minimum, faculty should retain self-archiving rights and rights for personal educational use and avoid granting an exclusive long-term license that extends beyond first publication.[18]

Thus, there is a need for faculty-authors to adopt an attitude towards copyright that is more sympathetic to their own non-commercial interests and to their primary educational purpose of advancing knowledge. However, the direct benefits of these changes need to be articulated and communicated to the faculty to actually effect change. These benefits include:

- Guaranteed freedom to use their own research material for teaching and other educational purposes. Faculty authors are sometimes unaware that transferring copyright can result in their inability to employ their own writings for teaching purposes, thus requiring them to seek permission before posting their work to their own web site or before using their work for course pack or library reserve purposes. When asked regarding various copyright and use issues, over half of faculty authors regarded the ability to use their own material for teaching (including course packs) to be important, and a third considered web-based self-archiving to be very important.[19] Therefore, this rights issue aligns well with faculty concerns.
- Increased flexibility as publishing channels and scholarly communications evolve. Faculty authors naturally tend to focus on current publishing media and channels when granting publishing rights. Equally naturally, publishers wish to gain rights broad enough to cover both current publishing channels and those yet to be discovered. Currently, scholarly journals have a virtual monopoly on conferring the prestige necessary for academic advancement. However, future venues may emerge that complement the journals' role. Limiting rights to first publication protects the author's ability to take advantage of such channels, without impairing the publisher's monopoly on first publication.
- Increased visibility. As we have discussed elsewhere, there is evidence that open access to research posted online increases the use and impact of the material. As this impact, both directly and indirectly, helps drive academic advancement decisions, it remains in the author's best interests.[20]

Obviously, individual faculty members—particularly junior faculty actively seeking tenure—are unlikely to withhold rights when they are demanded by a prestigious journal. Fortunately, the policies of an increasing number of academic publishers, especially

[18] For a fuller discussion, see Bennett (1999).
[19] See ALPSP (2002), pp. 23-24.
[20] See Lawrence (2001).

society publishers, reflect a genuine interest in accommodating author needs when those needs can be met without jeopardizing the publisher's legitimate business interests.[21]

Intellectual Property Issues & Author Rights

Resources & Further Reading

- The "Scholarly Electronic Publishing Resources: Legal" section of the "Scholarly Electronic Publishing Bibliography" provides links to many resources, including news, directories and guides, mailing lists and weblogs, organizations, publications, and U.S. laws pertaining to legal publishing issues, including copyright and author rights. Available from <http://info.lib.uh.edu/sepb/rlegal.htm>.

  Bailey, Charles W., Jr. *Scholarly Electronic Publishing Bibliography*. Houston: University of Houston Libraries, 1996-2002. Available from <http://info.lib.uh.edu/sepb/sepb.html>.

- Yale University Library's Liblicense provides a comprehensive guide to licensing issues including licensing terms; licensing vocabulary; model author and publisher licenses. Available from <http://www.library.yale.edu/~llicense/authors-licenses.shtml>.

- RoMEO (Rights Metadata for Open archiving), based at Loughborough University, is a project funded by the U.K. Joint Information Systems Committee (JISC) to investigate rights issues relevant to self-archiving of research at U.K. institutions of higher education. A specific goal will be the development of simple rights metadata that can be assigned to papers deposited in institutional archives and harvested via OAI-compliant service providers. See: <http://www.lboro.ac.uk/departments/ls/disresearch/romeo/index.html>.

- American Association for the Advancement of Science report on intellectual property rights and digital dissemination. "Seizing the Moment: Scientists' Authorship Rights in the Digital Age" calls for authors as the creators of scientific content to negotiate license agreements with scientific publishers that will maximize access to their work. The report is available from <http://www.aaas.org/spp/sfrl/projects/epub/epub.htm>.

- Scott Bennett. "Authors' Rights." *Journal of Electronic Publishing* Volume 5, Issue 2 (December 1999).

- Scott Bennett. "Position Paper on Yale University Copyright Policy." (March 1998). Available from <http://www.library.yale.edu/~llicense/bennett.html>.

- Mary M. Case and Prudence S. Adler. "Promoting Open Access: Developing New Strategies for Managing Copyright and Intellectual Property." *ARL Bimonthly Report* 220 (February 2002). Available from <http://www.arl.org/newsltr/220/access.html>.

- Rob Kling and Geoffrey McKim. "Not Just a Matter of Time: Field Differences and the Shaping of Electronic Media in Supporting Scientific Communication." *Journal of the American Society for Information Science*. Volume 51, Number 14 (2000): 1306-1320.

Undermining the Existing Journal Publishing System

Another concern that faculty share with academic administrators—one inherent in many of the faculty apprehensions discussed above—is that institutional repositories will undermine the current system of scholarly journal publishing. The scholarly journal publishing system serves an important role for faculty-authors in many disciplines: in addition to editorial quality control and dissemination, journals provide the legitimacy and prestige that drive professional advancement. Therefore, in presenting institutional repositories to faculty, one must bear in mind that faculty-authors (many of whom are

[21] The American Physical Society's publishing agreement provides just one example (<http://forms.aps.org/author/copytrnsfr.pdf>).

also journal editors and reviewers) are frequently sympathetic to the role played by scholarly publishers, particularly as the agents of peer review and quality control.

Institutional repositories will not and cannot, by themselves, eliminate the roles currently served by scholarly publishers, nor should they aspire to do so. One can project scenarios, as we have done elsewhere,[22] wherein institutional repositories provide a critical component in an alternative system of scholarly communication and publishing that delivers considerable benefits for the practice and economics of scholarly communication. However, in the vast majority of cases, faculty will be unaware or even skeptical of this broader potential, and pointing to such future scenarios will only fire the imaginations of a small proportion of an institution's faculty-author stakeholders.

All this suggests better success when institutional repositories are recognized as complements to, rather than as replacements for, traditional fee-based journals.[23] This allows repository proponents to build a case for faculty participation based on the primary benefits that repositories deliver, rather than relying on secondary benefits and on altruistic faculty commitment to reforming a scholarly communications model that has served them well on a personal level.

Both formal and informal scholarly communication practices—such as sharing preprints, communicating conference proceedings, participating in online discussion lists, building shared disciplinary databases, building shared disciplinary resource compendia,[24] and receptivity to online-only journals—vary by discipline. And many of these discipline-specific resources complement, rather than compete, with traditional academic journals. Presenting institutional repositories as analogous to—and as dissemination channels for—these existing and often time-honored practices might help overcome faculty resistance in some disciplines.[25] Further, careful responses to the faculty concerns discussed above—for example, in protecting essential publisher rights, as well as author rights when discussing author-publisher rights agreements—will also help allay faculty fears by reinforcing the concept that institutional repositories can coexist with the existing journal publishing system.

Reassuring Faculty Regarding the Existing Journal Publishing System

Resources & Further Reading

- See the "I Worry About…" FAQs at the EPrints.org site for responses to faculty objections on a number of issues. Available from <http://www.eprints.org/self-faq/>.
- Rob Kling, Lisa Spector, and Geoff McKim. "Locally Controlled Scholarly Publishing via the Internet: The Guild Model." *CSI Working Paper* no. WP-02-01 (June 2002). Available from <http://www.press.umich.edu/jep/08-01/kling.html>.

Faculty Workload

Not surprisingly, the effort many faculty might be expected to make to participate in an institutional repository will correlate to the benefits they expect to derive from it. In the early stages, therefore, when such benefits are less well understood—and, in any case, exist primarily in prospect—resources might have to be committed to support faculty posting to the repository, thereby lowering both the perception and reality of the effort falling to the author-participant.

---

[22] See Crow (2002).

[23] See Pinfield, Gardner, and MacColl (2002) and Bentum (2001b).

[24] For example, molecular structures, genetic maps, cumulative bibliographies, core text corpora, etc.

[25] See Kling and McKim (2000).

Although the EPrints software, on which many early repository implementations are based, is associated with author self-archiving, self-posting through the system requires several steps that may dissuade new and intermittent contributors. Given the significant disparity of technical proficiency amongst faculty, potential contributors might not have the expertise—nor the inclination—to deposit materials themselves.

Not surprisingly, then, early repository implementers consider library mediation of content submissions to be the only practical method of managing the archive, at least initially.[26] This library management of the document contribution process typically includes:

- Converting documents to allowed or preferred digital formats;
- Assigning metadata and subject headings and/or reviewing author-assigned metadata or headings;
- Providing faculty-authors with information regarding copyright and intellectual property issues. This can also involve providing information about the self-archiving policies of individual publishers, and even negotiating with individual publisher on behalf of contributing faculty; and
- Quality control and other ingest-related and administrative processes.

One way to ease and encourage faculty and departmental participation is to frame participation in a manner that it addresses a problem the faculty wishes to solve. By helping collect and host papers for a university-sponsored conference, assuming responsibility for departmental working paper series, or taking on digital production and archiving responsibility for existing programs, repository implementers can lessen the workload of faculty while actively encouraging their participation.[27] At the same time, such projects will have to be sensitive to the perceptions and apprehensions of the departmental support staff currently responsible for them. The user community orientation adopted by DSpace provides another alternative: each DSpace community designs a workflow process that accommodates the needs of its faculty and staff. In this way, administrative and technical responsibilities can be shared by the community's resources, coordinated with the library.[28]

Faculty Workload

Resources & Further Reading

- William J. Nixon. "The evolution of an institutional e-prints archive at the University of Glasgow." *Ariadne* 32 (2002). Available from <http://www.ariadne.ac.uk/issue32/eprint-archives/>.
- Pinfield, Stephen, Mike Gardner, and John MacColl. "Setting up an institutional e-print archive" *Ariadne* 31 (2002). Available from <http://www.ariadne.ac.uk /issue31/eprint-archives/intro.html>.

Impact of Discipline-specific Practices

In presenting the global potential of institutional repositories, we sometimes lapse into pan-disciplinary abstractions that imply that such repositories represent the logical convergence or homogenization of the scholarly communication needs of all academic

[26] See Pinfield, Gardner, MacColl (2002) and Nixon (2002).

[27] Caltech has relied almost exclusively on this approach to gain participation in its repository in the early stages. Personal communication, Kim Douglas, Caltech Libraries, September 25, 2002.

[28] Personal communication, MacKenzie Smith, MIT Library, October 30, 2002.

disciplines.[29] However, actual repository implementations must address goals at once more modest in scope, while more demanding in execution. While institutional repositories must participate in a global interoperable network to achieve their full potential, they must accommodate the varied needs of their local user bases. A global online network of interoperable research repositories will result from success at the local level adapting to the dynamic needs of specific user communities and the practical benefits they thus deliver to faculty authors and researchers.

Discipline-specific e-print servers have met their greatest success in those disciplines with existing prepublication traditions (for example, physics and mathematics). The narrow success to-date of discipline-specific e-print repositories demonstrates that digital publishing models that work well in one discipline will not necessarily translate well into other fields with more conservative practices for formal certification and quality indicators for research.

Some advocates of open access digital repositories consider them the most efficient means to communicate scholarly research and view their eventual adoption across academic disciplines as inevitable.[30] Others argue that heterogeneous discipline-specific publishing and communication practices are "durable in the medium-term," and that it is not just a matter of time before various academic disciplines converge on common digital communication channels to support scholarly communication.[31]

In either event, the evolution of practice in disseminating research will almost certainly have to come from within each academic community, rather than be imposed from outside. Still, institutional repositories, increased exposure to new distribution technologies, the practices of other disciplines, and the infusion of a new generation of scholars might well accelerate the rate of change in many fields. In the meantime, institutional repository implementations need to accommodate the various academic sub-cultures of an institution's schools or divisions. This accommodation will probably not be achieved by identifying common practices across disciplines and designing systems with universal applicability, but by providing the various disciplines with sufficient flexibility and autonomy to participate in the repository on their own terms. Individual communities of users can then set their own content policies and submission guidelines, within very broad limits, in such a way to encourage repository participation.[32]

Discipline-specific Differences

Resources & Further Reading

- Rob Kling and Geoffrey McKim. "Not Just a Matter of Time: Field Differences and the Shaping of Electronic Media in Supporting Scientific Communication." *Journal of the American Society for Information Science*. Volume 51, Number 14: 1306-1320 (2000). Available from <http://arxiv.org/abs/cs.CY/9909008>.
- Kling, Rob, Lisa Spector, and Geoffrey McKim. 2002. "Locally Controlled Scholarly Publishing via the Internet: The Guild Model." CSI Working Paper no. WP-02-01 (June 2002). Available from <http://www.press.umich.edu/jep/08-01/kling.html>.

---

[29] Kling and Lamb (1996) have demonstrated the failure rate of utopian technological visions that fail to adequately address the complex social realities of the intended adopters.

[30] See, for example, Ginsparg (2001).

[31] See, for example, Kling and McKim (2000).

[32] Kling and McKim (2000) provide a instructive look at the heterogeneous scholarly communication channels of various disciplines and the implications of digital media for the evolution of these channels. The design philosophy of the DSpace system is based on such a discipline- and community-specific focus.

- Rob Kling and Geoffrey McKim. Scholarly communication and the continuum of electronic publishing. *Journal of the American Society for Information Science* Volume 50 (1999): 890--906.
- The ALPSP (2002) survey takes into account attitudes and perceptions across academic disciplines, as does the ARNO survey (Bentum 2001), albeit with a small sample. More work needs to be done and shared to help repository implementers tailor their services to the practices of individual disciplines.
- Stephen Pinfield. "How do Physicists Use an E-Print Archive?" *D-Lib Magazine* Volume 7, Number 12 (December 2001). Available from <http://www.dlib.org/dlib/december01/pinfield/12pinfield.html>

  Provides insight into lessons learned from the arXiv high-energy physics e-prints server and their practical application to a multi-disciplinary institutional repository at the University of Nottingham. Includes an analysis of author experiences with self-archiving at the arXiv e-print archive <www.arxiv.org>.
- Paul Ginsparg. "Creating a global knowledge network." Invited contribution for Conference held at UNESCO HQ, Paris, February 19-23, 2001, Second Joint ICSU Press - UNESCO Expert Conference on Electronic Publishing in Science, during session *Responses from the scientific community*. Available from <http://arXiv.org/blurb/pg01unesco.html>.

Other Authors and Contributors: Students and Non-faculty Researchers

The above examination focuses on the interests and concerns of faculty authors, whose works typically represent an institutional repository's critical mass of intellectual output. However, there are, of course, other populations within the institution—including students and non-faculty researchers—whose works may be highly relevant and valuable to the repository program, if not crucial to its success. Staff researchers will frequently share the concerns of faculty authors and may be best addressed together with faculty (including the matter of voluntary participation). Student authors are potentially pre-disposed to the prestige and recognition, and their own form of academic advancement, that postings in the repository would present. Unlike faculty, to whom impositions of formatting standards and submission requirements would be problematic, one suspects there will be no such problems regarding students. Institutions typically prescribe rigid document format requirements for theses and dissertations, and students are accustomed to adhering to them. While one might anticipate students to adapt to digital publishing opportunities faster and with fewer reservations than faculty, graduate students will often be guided in such decisions by their faculty advisors, who might advocate a more conservative publishing approach.

LIBRARIANS: BENEFITS AND CHALLENGES

Libraries often provide the ideal institutional focus for these changes, as faculty often seem less skeptical of the library's motives than they are to those of their institution's administration. We noted above the heterogeneous scholarly communications needs specific to each academic discipline. Libraries and librarians can play a critical role in helping to facilitate the development of such digital communication channels tailored to the needs of individual disciplines. By providing the context and structure for the development of such channels through institutional repositories, librarians can apply their special skills and perspectives, as well as make effective use of the substantial resources being committed to research and communications by academic institutions, departments, government agencies, and individual researchers. Lack of such a coherent approach could result not only in the inefficient application of effort and resources, but in digital scholarly resources fragmented and effectively lost in marginal or moribund systems or repositories.[33]

[33] See Ginsparg (2001) and Kling and McKim (2000).

Thus, by driving and managing institutional repositories, libraries invest in the future and help maintain their relevance to faculty and administrators as digital publishing technologies and ubiquitous networking impact the structure of scholarly communication. Institutional repositories provide a mechanism through which librarians can work with faculty across disciplines as informal and formal scholarly communications channels evolve. Further, in this way, libraries can change their self-perception from being passive victims of perceived publisher malevolence to active agents for—and proponents of—their own relevance.

This implies expanded responsibilities and skill sets, although many of those required may already be the provenance of the library staff. Many aspects of the repository content ingest and administrative roles discussed below represent areas already familiar to librarians. This presents an opportunity for librarians to play a greater role in some scholarly communication functions—for example, registration and awareness—than they have in the past. For other functions, such as archiving, institutional repositories allow librarians to extend their traditional responsibilities to new media and new publishing models.

Librarians: Benefits & Challenges

Resources & Further Reading

- William J. Nixon. "The evolution of an institutional e-prints archive at the University of Glasgow." *Ariadne* 32 (July 8, 2002). Available at: <http://www.ariadne.ac.uk/issue32/eprint-archives/>

  Article recounts the experiences of the University of Glasgow in setting up an e-prints.org software repository (http://eprints.lib.gla.ac.uk). The article focuses on the practical implementation of the repository and the various decisions addressed in the course of the implementation.

- Stephen Pinfield, Mike Gardner, and John MacColl. "Setting up an institutional e-print archive." *Ariadne* 31 (April 11, 2002). Available from <http://www.ariadne.ac.uk /issue31/eprint-archives/intro.html>.

- The TARDIS Project (Targeting Academic Research for Deposit and dISclosure), sponsored by the University of Southampton and funded by JISC in the U.K., will examine ways to achieve the requisite cultural and institutional change necessary to encourage academics to self-archive. The project intends to investigate strategies for overcoming the technical, cultural, and academic barriers that currently impede the development of institutional e-print archives. See: <http://www.ecs.soton.ac.uk/~lac/TARDIS/bid.htm>

Encouraging Repository Participation

As a survey of early institutional repository implementations suggests, practical advocacy and education programs can assume a variety of forms.[34] These include:

- Producing a briefing paper for presenting the institutional repository case to relevant faculty and administration committees. This should be concise and include specific recommendations for action.

- Establishing a project web site (linked to/from the archive itself). This can act as a focus for developments and news.[35]

[34] See Nixon (2002) and Pinfield, Gardner, and MacColl (2002).

[35] See, for example, those for MIT's DSpace <http://www.dspace.org/>, Nottingham University <http://www-db.library.nottingham.ac.uk/ep1/information.html>, and Glasgow University(<http://www.gla.ac.uk/createchange/>).

- Identifying existing problems that the repository can solve for departments and faculty. Positioning institutional repositories as solving existing problems (albeit opportunistically) provides a more straightforward approach to encourage early participation than the presentation of more abstract, prospective benefits.[36]
- Presenting at departmental meetings and university committees.
- Distributing literature, such as the *Create Change* leaflet.[37]
- Placing articles, public service announcements, and advertisements in university magazines, the library user newsletter, and the like.
- Identifying champions amongst the faculty, particularly non-polarizing opinion leaders, to proselytize on the library's behalf.
- Developing an early adopter program with departments, labs, schools, university presses, and other entities that are likely to see the benefits of participation.[38]

Demonstration Programs

Achieving critical mass in terms of content is critical both to individual repository implementations, as well as to an interoperable network of online open access repositories. At the same time, gaining this critical mass requires that potential contributors understand the benefit that they might gain from participating in, and having access to, such a channel. This situation would pose a potential non-starter without a concerted effort to communicate and market the direct and secondary benefits that faculty-authors in particular would enjoy from such repositories. As we have discussed above, gaining faculty support and participation presents both the most important and most difficult aspect of implementing a repository.

The practical experiences of early repository initiatives suggest that the drive to gain content may be divided into two phases. In an initial short-term phase, repository sponsors gather sufficient content to demonstrate the potential and capabilities of the repository to potential contributors. In the second, long-term phase the repository achieves critical mass sufficient to provide a useful scholarly communication channel.

To assemble content for the demonstration program, repository administrators can locate research by faculty at their institutions that has already been posted to a discipline-specific server (for example, arXiv) or to personal or departmental web pages. The repository administrators can solicit permission from the authors to include the articles in the institutional archive, and may even discover additional research that can be posted. The demonstration site builds awareness and interest while serving as a facility for stakeholder feedback. This engages them in the development process that can lead to a full-scale repository program.

REPOSITORY MANAGEMENT AND POLICY ISSUES

We considered above potential faculty reservations about participating in institutional repositories, some of which vary by discipline. To aid repository implementers in formulating both content policies and practical content acquisition programs, we will now review specific types of potential content and the issues each type may raise.

[36] Personal communication from Kim Douglas, Caltech Libraries, September 27, 2002.

[37] Available from <http://www.createchange.org/change.html>.

[38] See, for example, DSpace's program: <http://www.dspace.org/join_us/lead_users.html>.

For practical reasons, we will focus here on content that supports the definition of institutional repositories provided above. Following this definition, content would typically be:

- Scholarly—the material is research- or teaching-oriented;
- Produced, submitted, or sponsored by an institution's faculty (and, optionally, students) or other authorized agent;
- Non-ephemeral—the work must be in a complete form, ready for dissemination;
- Licensable in perpetuity—the author must be able and willing to grant the institution the right to preserve and distribute the work via the repository.

Materials that satisfy the above requirements might include working papers; conference presentations; monographs; course materials; annotated series of images; audio and video clips; published (or pre-published) peer-reviewed research papers; and supporting material for published or unpublished papers (for example, datasets, models, and simulations).

While repository content may thus be defined broadly, some repositories may elect to focus initially on text-based materials, even though they anticipate broadening coverage over time. Additionally, in the interest of encouraging participation and acquiring material to populate pilot and demonstration projects, some repositories may choose to adopt more relaxed (and possibly temporary) guidelines for content in the repository's initial stages.

Repository Content: Published Material

Scholars in disciplines with no prepublication tradition will have to be persuaded to provide a prepublication version; as noted above, they might fear plagiarism or anticipate copyright or other journal acceptance problems in the event they were to also submit the work for formal publication. They might also fear the potential for criticism of work not yet benefiting from peer review and editing. For these non-preprint disciplines, a focus on capturing faculty postpublication contributions may prove a more practical initial strategy, addressing objections to repository participation.[39]

Including published material (or "postprints") will raise its own set of intellectual property issues, some similar to those for preprints. Given the faculty-author (and university administrator) attitudes and perceptions regarding the perpetuation of the existing scholarly journal publishing system and its relation to career advancement, publisher permissions and agreements become a critical factor in faculty repository participation.

As noted above, an increasing number of scholarly publishers—especially learned societies—are beginning to recognize that repository posting will not jeopardize the prestige, impact or economic health of their journals. Where a journal's author-publisher agreement does not grant such rights, institutions can negotiate with those publishers to allow embargoed (time-delayed) access to published research. Such embargoes would be based on the fact that readings of research articles—and hence, presumably, their economic value to the original publisher—drop precipitously one year after publication.[40]

---

[39] See Pinfield, Gardner, and MacColl (2002) and Bentum, Brandsma, Place, and Roes (2001).

[40] Tenopir and King (2000) show that amongst university scientists, over 58% of articles read are less than one year old, and over 70% are less than two years old. (See p. 189, Table 25). (The proportion of readings of newer material is higher for non-university scientists.)

While this pattern only reflects the reading of STM articles, one might anticipate a similar use decay curve for the social sciences and humanities.

When building content for a repository demonstration program or pilot project, implementers can mine published material from faculty and departmental web sites. Sometimes faculty will have the rights to post this published material; other times, one suspects, not. In any event, the process of securing the repository participation permission (discussed below) should help determine whether the author indeed holds such rights. In those instances where the author is unaware of, or indifferent to, the need to obtain such rights, the repository implementer can work with, or on behalf of, the faculty-author to secure and/or negotiate the necessary rights to post to the article to the repository—even after the article has been accepted under an existing author-publisher agreement.

Repository Content: Gray Literature

While surveys of faculty attitudes and perceptions reflect faculty-author concerns for the perpetuation of traditional scholarly journal publishing, they also indicate that faculty consider institutional repositories to be particularly well-suited for various types of gray literature and other fugitive and unpublished material.[41] This material includes:

- preprints;
- working papers;
- theses and dissertations;
- research and technical reports;
- conference proceedings;
- departmental and research center newsletters and bulletins;
- papers in support of grant applications;
- status reports to funding agencies;
- committee reports and memoranda;
- statistical reports;
- technical documentation; and
- surveys.

Such gray literature forms a part of the informal scholarly communication process we have discussed above. In some instances, an item may be followed by a formal publication. Often, however, that it is not the case and the material becomes difficult to identify and access, let alone preserve. Further, even when gray documents are subsequently published, significant detail—for example, on research methods and experiment techniques—is frequently omitted. Thus, while peer-reviewed journals provide the principal venues for formal communication within scholarly communities, informal gray literature serves a valuable supplementary role.[42] We will review some of the major types of gray literature below.

Preprints

Preprints serve two basic purposes:

- They establish intellectual priority in fast moving fields. In some scientific fields, the journal publishing cycle is too slow, or circulation too narrow, to provide the sole channel for disseminating research results and claiming priority. Additionally,

[41] See Bentum (2000b).
[42] See Weintraub (n.d.).

preprints help eliminate duplicative research by making researchers aware of the research activities of others.

- They attract critical response and comment that allows the paper to be refined and revised for formal publication in a journal.

Some disciplines have long-standing prepublication practices, with paper mechanisms predating digital implementations. High-energy physicists, for example, had a preprint culture that predated the application of digital media to the purpose, and ArXiv,[43] an e-prints server for high-energy physics, was originally designed to automate and improve this existing paper-based process.

In addition to ArXiv, other academic disciplines with established preprint traditions developed electronic mechanisms to facilitate the sharing and storage of research preprints. Discipline-specific digital repositories for economics (RePEc);[44] cognitive science (CogPrints);[45] astronomy, astrophysics, and geophysics (NTRS and ADS);[46] and computer science (NCSTRL)[47] evolved within those specific research communities as digital extensions of existing peer-to-peer research communication practice.

While the fields of management, business, and finance circulate working papers in a manner analogous to preprints, the RePEC economics e-print server has not achieved the same level of participation as arXiv. One reason may lie in the fact that many business schools/institutions publish such working papers as a series, providing another channel for preprint dissemination. Other fields have more restricted preprint cultures. Molecular biologists, for example, typically circulate preprints within small invisible colleges, with broader distribution depending on publication in scholarly journals. While several biological science e-print servers have been established, such servers do not play the central role as they do for high-energy physics.[48] In medicine, the posting of prepublication working papers is even considered as a danger to public health, if they are used as the basis for clinical practice or promulgated by the media.[49]

Recognizing and accommodating discipline-specific practices will enable an institutional repository to better anticipate and serve the needs of potential faculty contributors. Where a school or division charges for working paper series and generates an income surplus, for example, the institutional repository might have to restrict access to the material or allow an embargo to gain the content for the repository. On the other hand, where a working paper series charges solely to offset the costs of print distribution, the institutional repository can provide an alternative distribution channel providing broader dissemination via open access.

Overall, participation in electronic preprint or post-print servers is not yet a common practice for most disciplines (physics and mathematics being the most notable exceptions).[50] The ALPSP survey indicates that only about a tenth of faculty authors

[43] <http://arxiv.org/>

[44] <http://netec.mcc.ac.uk/RePEc>

[45] <http://cogprints.soton.ac.uk>

[46] NASA Technical Reports Server (<techreports.larc.nasa.gov/cgi-bin/NTRS>) and the NASA Astrophysics Data System (<http://adswww.harvard.edu/>).

[47] National Computer Science technical Reference Library (<http://ncstrl.org>).

[48] Kling and McKim (2000).

[49] See Pinfield (2001).

[50] See ALPSP (2002), p.21.

deposited preprints, and almost one-third of those depositing preprints were in physics.[51] Still, as the PrePRINT Network suggests, the practice of preprint posting is broader than many realize.[52]

Including preprints in a repository will inevitably raise quality control questions. Given a repository's potential to increase the visibility and prestige of an institution, the institution has a vested interest in the quality of the content. As we will discuss below, several existing repository programs delegate this responsibility to the institutional communities (departments, research centers, labs, etc.) best positioned to determine appropriate accession guidelines for content in their areas. While such vetting does not substitute for peer review, it does provide the institution with some basic level of quality control. This issue should be explicitly addressed by the repository's content accession policy.

In addition to quality control issues, including preprints in an institutional repository will raise the following issues:

- The contribution of preprints will be limited, at least initially, to disciplines with established prepublication traditions. Preprints raise a welter of issues (including plagiarism, info abuse, etc.) for many the disciplines without prepublication traditions.
- Even for some disciplines with a prepublication tradition, preprints will raise contributor concerns regarding future journal publication. For example, some publishers—particularly in medicine—require that online preprints be withdrawn once the article is published. This requires that policies address both rights assignment issues, as well as the ability of authors to withdraw access rights.[53]
- Where both preprints and post-prints are included, the repository will need to ensure that each type of document is clearly labeled. This is necessary to distinguish between versions of the same work and to address contributor concerns that repository working papers might give a partial view of their research.
- Sometimes an author will want to withdraw the preprint, either to satisfy a publisher or to avoid the impression that the preprint represents the latest state of the research. Obviously, this contingency conflicts with the repository's goal to maintain content in perpetuity. To resolve such potential conflicts, a repository's rights management policies and technical systems must take them into account.[54]

None of this is to suggest that an institutional repository should avoid the inclusion of preprints. Indeed, preprints can constitute one of a repository's most valuable content types. However, as the above indicates, besides establishing broad policies about the types of content it will include, an institutional repository must accommodate each discipline's existing peer-to-peer communication patterns and research practices when developing institutional repository content policies.

---

[51] See ALPSP (2002), pp. 13-14. Interestingly, an extensive survey of faculty perceptions and attitudes suggests that most faculty—even in scientific disciplines—have only a vague understanding of what e-print servers are.

[52] This site will help identify faculty members to serve as contributors to a repository pilot program; see <http://www.osti.gov/preprints/ppnbrowsec.html>.

[53] Respondents to the ARNO study (Bentum 2001a) indicated an interest in participating in the ARNO university server as long as they could withdraw documents at any time.

[54] See, for example, DSpace's policy statement on content withdrawal at <http://dspace.org/mit/policies/community-collection.html#withdrawal>.

Curriculum Support and Teaching Materials

Besides the benefits for faculty as authors, institutional repositories can also deliver benefits to teaching faculty. By including non-ephemeral faculty-produced teaching material, the repository serves as a resource supporting classroom teaching. These materials might include online lecture notes, concept illustrations, visualizations, models, simulations, course videos, and the like—much of the material often found on course web sites. This benefit should help extend the appeal of institutional repositories across a broader audience of research and teaching faculty. Including this material should also encourage broader participation in the repository, even by faculty-authors yet to be convinced of the merits of posting working papers or published articles.

Electronic Theses and Dissertations

Student electronic theses and dissertations ("ETDs") also provide logical content to be captured by institutional repositories, and to that extent, students are also author stakeholders in such repositories. Universities typically have comprehensively prescribed and meticulously enforced document format requirements for graduate dissertations. However, practical experience with electronic theses—including submission standards and requirements—varies with institution and in many instances such policies are still evolving.[55] Some repositories will elect to integrate access to student theses and dissertations with the Networked Digital Library of Theses and Dissertations, while others will maintain ETD material locally.[56]

Institutional Repository Content Issues

Resources & Further Reading

Gray Literature

- The European Association for Grey Literature Exploitation (EAGLE) in Europe is a co-operative network for identification, location and supply of gray literature. EAGLE is a non-profit association formed by the National Centres participating in SIGLE (System of Information for Grey Literature in Europe). See: <http://www.kb.nl/infolev/eagle/mission.htm>.
- The New York Academy of Medicine Library <http://www.nyam.org/library/index.shtml> maintains information and resources on gray literature and its importance to communicating scientific knowledge. See <http://www.nyam.org/library/greylit/index.shtml >.
- Irwin Weintraub. "The Role of Grey Literature in the Sciences." Available at: <http://library.brooklyn.cuny.edu/access/greyliter.htm>.

Benefits to Students

- The Networked Digital Library of Theses and Dissertations, under the auspices of Virginia Tech, provides a wealth of information on electronic theses and dissertations: <http://www.ndltd.org>.
- Gail McMillan, Edward A. Fox, and John L. Eaton (1999) "The Evolving Genre of Electronic Theses and Dissertations." *1999 Hawaii International Conference on System Sciences*.
- Thomas H. Teper and Beth Kraemer (2002) "Long-term Retention of Electronic Theses and Dissertations." *College and Research Libraries* 63 (1): 61-72.

[55] See <http://library.caltech.edu/collections/etd/guidelines/bodyformat.html>.
[56] See McMillan, Fox, and Eaton (1999) and the Networked Digital Library of These and Dissertations: <http://www.ndltd.org/>.

Defining Repository Communities

We have emphasized above that a critical success driver for institutional repositories will be the extent to which the implementers understand and accommodate the informal and formal scholarly communication processes of academic disciplines and sub-disciplines. Practically, this translates into integrating academic communities into the structure of the repositories content, policy, and management structure.

This integration can be accomplished in a variety of ways: MIT's DSpace has integrated this community-orientation into the structure of its repository support system, both from a policy and system development perspective.[57] Other repository implementations, while recognizing the importance of discipline-specific practices, have approached the issue less formally. In either event, the manner and extent to which academic communities themselves participate in a repository's administration and management will effect content definition and acquisition policies, as well as the practical steps of content ingest.

At a broad level, an institutional repository serves affiliated users—for example, students, faculty, and staff at the institution—as well as a global universe of unaffiliated users. The latter would comprise any persons accessing the repository's content either directly, through search and retrieval mechanisms that the repository might incorporate into its implementation, or through OAI-compliant discovery services that harvest the repository's metadata and make it broadly accessible. Users at this level, lacking any further authorization by the repository, would have the shallowest access to repository content.

Affiliated users, on the other hand, might often have greater access to repository content, with the extent of the access being based on community-specific rights and access management policies.

User Groups & Communities

To take one example, DSpace delegates decisions concerning what may be deposited in the repository, as well as the policies that governs its use, to the various communities that comprise the repository. This distributed administration recognizes both the realities of managing a repository in a large academic university environment, as well as the discipline-specific needs of each community. To further facilitate management, these communities typically correspond to administrative entities within the institution (for example, a department, school, research center, or laboratory). Besides providing a practical mechanism to ensure that the repository is discipline-driven, DSpace defines user groups in order to implement specific system functionality. For example, authorization to edit a user group home page, add content to the user groups, and submit items through a user group's submission process are all managed at the user group level.[58]

Regardless of technical system infrastructure, policy-based processes will allow repositories to specify content deposit approval process for each community, administered by individuals from the relevant user community. The complexity and rigor of the approval process can also vary to serve the needs of each community. Some communities will allow registered users to post content without qualitative vetting. Others will invoke approval layers that determine the appropriateness of the material and apply quality control standards.

Proponents of the "guild" model assert that working paper and occasional paper series—an established component of current communication practice for many disciplines—provide a logical model to extend online open access publishing

[57] See <http://dspace.org/mit/policies/index.html>.
[58] See Bass *et al* (2002).

incrementally across disciplines.[59] Sometimes the administering research community will already have such a working paper series in place, and repository participation standards for the community will coincide with existing standards and policies. In other instances, a department's or research center's selectivity in its hiring standards will lend legitimacy to the contributions of its members and serve as a strong quality indicator for papers submitted to the repository. These repository contributions would thus constitute a *de facto* occasional paper series, with a perceived quality between peer-reviewed contributions and the posting of unvetted preprints.

The guild model does not presuppose the existence of institutional repositories, but such repositories would provide a logical institutional and technical framework for guild-sponsored working paper series. Further, community-sponsored working paper series can be implemented locally, within the framework of an institutional repository, without requiring global, discipline-wide adoption of the model. Applied in the content of institutional repositories, the guild model can thus help advance faculty-author participation in institutional repositories.

Content Deposit Processes

Repositories may be set up to accommodate user communities, collections, or both. Existing repository system software allows different classes of users and digital resource collections according to resource type. Sometimes a community will comprise more than one group of users and more than one content collections. Collections typically comprise items that share one or more characteristics (for example, by purpose, source, subject matter, or audience). In this way, each collection can have its own content submission and approval process, as well as its own set of administrators and managers.[60]

Typically, an item submitted to a repository undergoes editorial and quality control reviews—the rigor of which vary from institution to institution and even between user communities within an institution—before being made publicly available through the repository. Depending on the system infrastructure, many of these review criteria can be automated (for example, cross-checking that the submitting author is approved to submit to a particular repository community or sub-repository), while others (for example, metadata review and augmentation) typically require manual intervention. The same basic document workflow applies regardless of the repository software infrastructure being used, with an item moving through various stages of initial deposit; review, correction, augmentation; and rejection/approval. A more detailed content deposit workflow is described below:

- Author or author proxy submits an item to the repository.
- The author accepts (or rejects) a permission agreement that grants the host institution sufficient rights to make the item available to end-users and to convert it as necessary for digital access and preservation purposes.
- A review determines that the submitter is authorized to contribute to the repository (or sub-repository) to which he or she has submitted the item. This review enforces the institution's repository policies regarding the submitting author's institutional affiliation and status (for example, faculty, staff, student), the subject area of the item, community-specific approval processes, and other selection criteria established by each repository.
- A review verifies, and augments if appropriate, the metadata submitted with the item.

[59] See Kling, Spector, and McKim (2002).
[60] See Gutteridge (2002) and Bass *et al* (2002).

  - This metadata makes it possible for users to search and/or browse to find the item and for internal management of the repository content.
  - Most repositories will support some baseline level of metadata (typically based on the Dublin Core), while others will also support domain-specific metadata.

- A review determines whether the submitted item is in a known and/or approved document format.
  - This ensures that the item will be readable to those users who have access to it, and may allow for it to be converted to a supported format type; and
  - This also supports the archival preservation of the item by allowing management of document format types and the migration of formats at some subsequent stage.
- At any of these review stages, an item might be:
  - Rejected as inappropriate and deleted from the repository (for example, the author is not authorized to submit to the repository;
  - returned, with comments, to the submitter for emendation and resubmission; or
  - accepted and posted to the repository.
- Once the item is accepted, it is assigned a unique document identifier and a persistent URL to ensure its perpetual availability.
  - By definition, institutional repositories intend to make submitted content available in perpetuity. Unique document identifiers allow the content to outlive the repository infrastructure itself.

Ideally, an institutional community can skip any or all of these steps of the content approval process, giving user communities flexibility in managing their collections.[61] From a functional perspective, the above workflow would typically include:

- reviewers—those who review the content to determine that it is appropriate for the collection to which it has been submitted;
- approvers—those who check the contribution for completeness and obvious errors. Sometimes the people who fulfill this function will also have editing rights, depending on the user community; and
- metadata editors—those who check and/or augment the contribution's metadata.[62]

Distribution Licenses

To allow the host institution to administer and disseminate the material submitted to the repository, the repository will need each contributor to grant the institution an irrevocable, non-exclusive, royalty-free license to distribute the content, to translate its format for the purpose of digital preservation, and to maintain the content in perpetuity.

---

[61] See, for example, Bass *et al* (2002).

[62] DSpace has settled on four functions: "Submitter," "Content Reviewer," "Metadata Editor," and "Coordinator." DSpace implementers opted for the more neutral "coordinator" over "approver" after they encountered resistance to the idea of someone outside a community "approving" content. Personal communication, MacKenzie Smith, MIT Libraries, October 30, 2002.

Theoretically, such license agreements might vary by user community and/or by the type of content collection, with implications for the rights management mechanisms we will discuss below.[63]

Defining Repository Communities

Resources & Further Reading

- Kling, Rob, Spector, Lisa, and McKim, Geoff. "Locally Controlled Scholarly Publishing via the Internet: The Guild Model." CSI Working Paper no. WP-02-01 (June 2002).
- See Bass, Michael J. *et al*. DSpace: Internal Reference Specification: Technology and Architecture. Version 2002-03-01 (2002). Available from <http://dspace.org/technology/architecture.pdf>.
- For an overview of Caltech's open access digital archives, see: <http://coda.caltech.edu>.
- For DSpace's author permission agreement, see: <http://dspace.org/mit/policies/license.html>.
- For Caltech's sample author permission agreement, see: <http://resolver.caltech.edu/caltechLIB:2001.002>.

TECHNICAL & SYSTEM ISSUES

Addressing the many and varied issues discussed above will prove essential to implementing an institutional repository, as well as to reaching out to faculty authors to secure their participation. At the same time, the repository requires a technical infrastructure that supports the repository's goals of preserving an institution's intellectual output, while making the content broadly available through interoperability with other open access repositories. This technical implementation could be quite simple: a hierarchical file structure, web access, and OAI-compliant metadata would allow users to employ OAI search engines in finding and retrieving repository content.

Fortunately, however, repository system solutions exist that will serve the needs of the vast majority of institutional contexts, while providing a wide range of administrative and end-user features and functionality. Evaluating the suitability of these solutions for a particular implementation, and making specific implementation decisions, requires an understanding of the basic technical issues, initiatives, standards, and protocols that support essential repository functionality. To support this evaluation, we provide overviews of these basic concepts and initiatives below.

While several initiatives are developing system infrastructures that support institutional repository implementations, the two most widely discussed systems are the Eprints software and the DSpace system. Developed at the University of Southampton,[64] the EPrints software is, by all accounts, relatively easy to install and configure to suit an institution's requirements, although it does require some proficiency with MySQL and the Perl scripting language. The software, which is open source, requires the Linux operating system,[65] the Apache web server, the MySQL relational database management system, and the Perl module.[66]

---

[63] To accommodate the reality of license terms changing over time, DSpace stores a copy of the license granted the submission of the item with item itself, making the specific license terms for any item always available. See Bass *et al* (2002).

[64] <http://www.eprints.org>.

[65] Though designed to run under GNU/Linux, EPrints has also been reported to run under other versions of Linux as well. There are no plans for a Windows version of the system. See Gutteridge (2002), p.8.

[66] See Gutteridge (2002), p.9.

In mid-2002, the University of Southampton established a strategic partnership with Ingenta PLC. This partnership is intended to allow Ingenta to use Southampton's EPrints software as part of a planned suite of OAI-related services, potentially including a commercial OAI-compliant hosting service that would serve institutions that elect to outsource their repositories. Ingenta has also indicated that it will feed any enhancements that it makes to the EPrints platform back into the EPrints/OAI community.[67]

DSpace, a collaborative project of the MIT Libraries and the Hewlett-Packard Company, has created a repository system that can support a federation of institutional repositories.[68] Because of its focus on the specific requirements of the institutional repository, DSpace design and functionality pays particular attention to the content input side of the process. The system was also designed to integrate with third-party software, allowing it to be coupled with other components (for example, editorial workflow systems) to render a turnkey publishing system. The DSpace code will eventually be released as Open Source.[69]

Development & Operational Costs

As with most any technology-based enterprise, one generally thinks of expenses in five categories: labor (and the equivalent if some skill requirements are met via out-sourcing), software, hardware, network, and depending on institution practices, overhead.

The technical support costs of developing and operating an institutional repository will depend on the service level agreement the repository has with the institution's technical support operations, and possibly, with third parties. Implementers of EPrints software indicate that the staff time required to install and configure the software is approximately four to five FTE days. While other library staff can perform much of the policy-based component of the repository, setting up the repository technical infrastructure—even using a largely turn-key solution such as the EPrints software—requires the assistance of a technical systems administrator.[70]

Software costs will depend on a basic "build or buy" (or "borrow") decision, which has economic, strategic, and many practical considerations. As discussed elsewhere, a number of proven, dependable, flexible, low-cost software solutions are available. "Buy" implies a level of effort over an extended time that will deter most new institutional repository implementers.

Hardware costs depend on the performance, storage, and other attributes of the configuration selected. EPrints can run on a basic hardware configuration, although disk storage, server capacity, and perhaps other specifications would need to be upgraded as the repository moved from a pilot stage into public operation and heavy use.[71] Hardware specifications for DSpace are not yet available. However, system hardware costs for either system will vary with the fault tolerance that the repository is willing to accept (for

---

[67] See: <http://www.ingenta.com/isis/general/Jsp/ingenta?target=/about_ingenta/press_releases/southampton.jsp>.

[68] <http://dspace.org/index.html>.

[69] See <http://www.dspace.org/live/home.html>.

[70] Informal estimates place the level of IT effort at half an FTE staff position for an experienced systems administrator. (Personal communication, Kim Douglas, Caltech Libraries, September 27, 2002.) However, after initial set up the process tends to require sporadic attention rather than full-time staff support. (Personal communication, Chris Gutteridge, University of Southampton, October 14, 2002.)

[71] System hardware with the general specifications cited by Pinfield, Gardner, and MacColl (2002)—Intel Pentium iii processor; 800 MHz processor speed; 256MB RAM; 20GB IDE disk—would cost approximately US$2,000 at this writing.

example, low downtime tolerance might require an inventory of replacement drives, etc.), backup capabilities, and other requirements. The cost of such services will typically depend on the existing capabilities of such units and the extent to which the repository implementation can achieve operating efficiencies with existing technical operations. The same is true of networking, which should be a modest incremental expense to the institution's existing network.

Non-technical labor costs, including user support, marketing and advocacy, and program administration, will typically outweigh the requirements for technology staff. On-going technology labor costs, such as for system administration, are generally allocated as an increment of existing human resources and programs. Initially, non-technical staffing may also be handled via resource allocation, although larger initiatives will need to commit to staffing long-term program management positions.

Finally, overhead costs may or may not be material, depending upon the institution's practices. Obviously, proponents of the new institutional repository will need to present a full budget and probably multi-year forecasts at some point in their interaction with university and library administration.

The Ability to Migrate and Survive

When considering a technical implementation for an institutional repository, it is important to remember that the explicit expectation is that the content managed by the system will survive the system itself and can migrate as new technologies evolve. Therefore, the system must be content-centric: applying standards and protocols that facilitate ongoing access to the information itself must be central to the system's conception. The design and implementation of both the EPrints software and the DSpace system have been based on such standards. EPrints can export the archive metadata in XML in a structured format that facilitates migrating to a subsequent system.[72] Both EPrints and DSpace are based on open source software licensing principles.[73]

In any event, switching costs from one repository technical solution to another would typically be high. Also, switching systems and solutions can be quite risky. Therefore, institutions will want to select their implementation path carefully. Even though several of the solutions are open source, they still involve database mapping and other customizations that would require additional investment if the infrastructure were changed.

EPrints and DSpace offer off-the-shelf systems that allow an institution to implement a complete framework for an OAI-compliant repository without resorting to in-house technical development. Both systems can be customized to meet local requirements, allowing an institution to configure metadata formats, design subject hierarchies, define acceptable file formats, and register with OAI.[74]

---

[72] Personal communication, Chris Gutteridge, University of Southampton, October 14, 2002.

[73] The operating system and all of the supporting software for EPrints are Open Source software licensed under the GNU General Public License (GPL). (See <http://www.fsf.org/copyleft/gpl.html> and <http://www.eprints.org/download.php> for full details.) MIT and Hewlett-Packard have agreed to license all DSpace software with an open source, BSD license. See Bass *et al* (2002). DSpace intends to add any third-party components under the same terms.

[74] Pinfield, Gardner, and MacColl (2002.); Bass *et al* (2002); and Gutteridge (2002). Additionally, EPrints supports multilingual implementations. (Personal communication, Chris Gutteridge, University of Southampton, October 14, 2002.) For an example of a multilingual implementation see: <http://papyrus.bib.umontreal.ca> which operates in both French and English.

Technical System Issues

Resources & Further Reading

Institutional Repository System Overviews

- Christopher Gutteridge and Stevan Harnad. "Applications, Potential Problems and a Suggested Policy for Institutional E-Print Archives." (August 19, 2002). Available from: <http://eprints.ecs.soton.ac.uk/archive/00006768/>
  The University of Southampton has been running a digital publications archive since 1998. This article provides practical implementation insight and advice from both the policy and technical perspectives.
- DSpace Technical Architecture Specification Document. See: <http://web.mit.edu/dspace/live/implementation/design_documents/architecture.pdf>
- DSpace Functionality Specification Document. See: <http://web.mit.edu/dspace/live/implementation/design_documents/functionality.pdf>.
- The University of Rochester's analysis of potential technology solutions relevant for an institutional repository implementation will provide a useful, brief overview for institutions just beginning to explore system options. Susan Gibbons. "Seeking a System for Community-Driven Digital Collections at the University of Rochester." *SPARC E-News* (February-March 2002). Available at: <http://www.arl.org/sparc/core/index.asp?page=g23#5>.

EPrints Software Implementation Descriptions

Several articles detail institutions' experiences implementing the EPrints software:

- Pinfield, Stephen, Mike Gardner, and John MacColl. "Setting up an institutional e-print archive" *Ariadne* Issue 31 (2002). Available from <http://www.ariadne.ac.uk /issue31/eprint-archives/intro.html>.
  Article outlines the major issues involved in establishing an institutional repository based on the experiences of the universities of Edinburgh and Nottingham.
- William J Nixon. "The evolution of an institutional e-prints archive at the University of Glasgow" *Ariadne* Issue 32 (2002). Available from <http://www.ariadne.ac.uk /issue32/eprint-archives/intro.html>.

  And:
- Chris Rusbridge and William J. Nixon. "Setting up an institutional ePrints archive—what is involved?" Unpublished paper, UKOLN Meeting (July 11, 2001). Available from <http://www.lib.gla.ac.uk/eprintsglasgow.html>.
  Both articles describe the implementation experience of the University of Glasgow.
- Sponsler, Ed, Van de Velde, Eric F. "Eprints.org Software: A Review." *SPARC E-News* (August-September 2001).
  A review of the EPrints software (version one) based on Caltech's experiences implementing a pilot institutional repository. The Caltech implementation includes multiple content repositories, including several technical-report repositories and one online conference proceedings. Our repositories are available at (for the Caltech digital repositories, see: <http://coda.caltech.edu>).
- Ed Sponsler. "Eprints from Scratch: A step-by-step guide to creating an electronic archive of scholarly documents." (forthcoming).
  This guide, by the IT lead responsible for establishing several OAI-compliant repositories at Caltech, provides a detailed "how-to" approach to setting up and maintaining an institutional repository. The guide includes explicit and lucid explanations of the installation and configuration of all the software--from the Linux operating system on up—required to support an EPrints-based system. While the discussion focuses on an e-prints.org system, many of the issues covered will prove relevant regardless of the system being implemented. The guide's intended audience includes IT specialists (of all

experience levels), as well as librarians and others who might benefit from an understanding of the technical mechanisms that support an institutional repository.

- Christopher Gutteridge. *EPrints 2.1 Documentation* (July 10, 2002). Available from <http://software.eprints.org/documentation.php>.
- The EPrints mailing list (<http://software.eprints.org/tech.php/>) provides an ongoing forum on EPrints software features, new capabilities, and support issues. Knowledgeable EPrints developers and staff answer questions and respond to questions about the software and relate issues.

Digital Content: Document Formats

As indicated above, an institutional repository might include a broad array of disparate document types. This suggests that the repository will also have to be able to accommodate a variety of digital file formats, including widely used formats such as ASCII, Postscript, Rich Text Format, and PDF. Additionally, content policy will have to determine whether the repository will accept other generic formats (for example, HTML), proprietary word processing formats (for example, Microsoft Word), and discipline-specific text editors (for example, TeX or LaTeX, used by mathematicians and physicists), images, and streaming media. Accepting some specialized formats might depend on the ready availability of translation programs to convert files from non-supported to supported formats. For example, open source utility programs exist to convert LaTex to Postscript or PDF.[75]

Precisely and rigidly dictating digital file formats for faculty will prove problematic, for both attitudinal and practical reasons. To simplify content deposit and encourage faculty participation, the institution will want to accommodate the wide range of file formats popular with various academic departments. At the same time, the repository needs to balance the desire to accommodate content contributors with the complications that migrating some of those formats or media might present as new standards evolve.

Besides file format, the repository will need to develop technical specifications for the repository's digital resources. This definition is both a policy and a technical issue. The EPrints software allows an implementing institution to specify the document types and formats that it will accept. It also allows the institution to identify content as "published," "in press," or "unpublished," providing the transparent content labeling identified above as critical to faculty acceptance.[76] Additionally, DSpace allows items to comprise multiple files. (For example, a conference paper along with the overhead presentation delivered at the conference; research papers and supporting datasets; etc.) DSpace intends to use the METS metadata standard to store the relationships between components in a bundle of items.[77]

Digital Content: Longevity

As the provision of long-term access and preservation are also essential elements of an institutional repository's mission, the need to preserve these multifarious digital objects must also be addressed. This is important both for the repository to preserve the intellectual product of a given institution, but also to form a component in an interoperable network of content repositories. Providing such long-term access to digital objects in the repository requires considerable planning and resource commitments.

The importance given to the preservation of repository content will vary with each institution. Some will assign considerable importance to such preservation from the

[75] See Pinfield, Gardner, and MacColl (2002).
[76] See Gutteridge (2002), p. 20 and Nixon (2002).
[77] See Bass *et al* (2002).

outset. Others will recognize the significance of the issue, but defer further attention until progress has been made in terms of developing standards for digital preservation. These issues will also be a function of the repository software system implemented.

When deciding which file formats to accept and maintain, the repository must address the issue of preserving the document in digital format. There are three strategies for long-term digital preservation:

- Preserving obsolete technologies: as this entails maintaining every version of every piece of hardware and software necessary to access the preserved data, it is not generally considered to be a viable alternative.
- Emulation: emulation essentially uses software to mimic the content's original software and hardware platform. In other words, the computer environment, rather than the data itself, evolves over time. While emulation is considered to hold considerable promise, it is largely unproven in digital preservation.
- Migrating digital content: this strategy involves the periodic transfer of digital content through successive hardware and software platforms. Migration requires a unique solution for each format that is to be converted. Some forms of migration are well established, and it is often regarded as the most promising method when data formats can be limited to standard formats. However, since the evolution of future formats will always remain unknown, and costs are recurring and unpredictable, it is difficult to predict the costs and efficacy of this approach.[78]

The Open Archival Information System (OAIS) Reference Model,[79] the *de facto* standard for digital archive architecture, provides the framework within which preservation metadata and other standards can be developed. The OAIS model is predicated on capturing content as bitstreams which can then be preserved in perpetuity.[80]

While many of the early institutional repository implementations have deferred decisions about long-term digital preservation, for preservation purposes the DSpace system captures the specific formats of files that users submit. DSpace maintains a bitstream format for each bitstream stored in the system. The system maintains a registry of known bitstream formats, and automatically identifies the format when possible. For unknown formats, the system queries the submitter requesting additional information. System administrators maintain the registry of known format types and the preservation service level available for each format type. However, where the format of the bitstream is unknown, the repository can make no claims regarding preservation and future use of the file.[81]

Preservation metadata provides the information infrastructure that supports the processes necessary to ensure that the bitstreams can be read, processed, and used over time. Such preservation metadata facilitates the management of a repository's content, compared to

---

[78] See "Technological Obsolescence" on the PADI website: <http://www.nla.gov.au/padi/topics/13.html>.

[79] OAIS should not be confused with the Open Archives Initiative (OAI). OAIS focuses on the preservation and archiving of digital objects, while OAI focuses on an explicit protocol for metadata harvesting that facilitates the interoperability of digital repositories. The OAI and OAIS have different, although orthogonal, goals relative to digital repositories. For more on OAI, see below.

[80] For more on the OAIS reference model, see: Lavoie (2000) and the OAIS resources listed in the Resources & Further Reading section.

[81] For more on DSpace preservation service levels, see Bass *et al* (2002).

descriptive metadata schemas (for example, the Dublin Core), which facilitate the discovery and identification of digital objects.[82]

Preservation Outsourcing

While some institutions will handle digital preservation locally, others will elect to manage the administrative, policy, and intellectual aspects of the repository, and contract with a trusted third party provider for the repository's digital file storage and maintenance. A recent OCLC/RLG report establishes a framework of attributes and responsibilities for sustainable digital repositories capable of handling large scale, heterogeneous collections of digital materials.

The OCLC/RLG framework identifies the attributes of a trusted digital repository important to institutions considering outsourcing the digital preservation function. These attributes include: standards compliance, administrative responsibility, organizational viability, financial sustainability, technological and procedural stability, system security, and procedural accountability.[83]

Scalability

The cumulative nature of institutional repositories also implies that the repository's infrastructure must be scaleable. As we have discussed, whatever a repository's content deposit criteria, items once deposited cannot be withdrawn—except in presumably rare cases involving allegations of libel, plagiarism, copyright infringement, or "bad science."[84] While initial processing and storage requirements might prove modest, institutional repository systems must be able to accommodate thousands of submissions per year, and eventually must be able to preserve millions of digital objects and many terabytes of data.[85] Further, storage requirements will depend on the formats the repository accepts. No reliable models yet exist to project data accretion rates and scale disk storage requirements for institutional repositories, although existing repository implementations are making an effort to develop such models.

Digital Content: Formats & Preservation

Resources & Further Reading

- "Trusted Digital Repositories: Attributes and Responsibilities." An OCLC-RLG Report. Research Libraries Group (May 2002).
  The OCLC/RLG report establishes a framework of attributes and responsibilities for sustainable digital repositories capable of handling large scale, heterogeneous collections of digital materials. The framework helps institutions faced with building local digital repositories or with identifying third parties capable of serving their digital preservation needs. See:
  <http://www.rlg.org/longterm/repositories.pdf>

[82] This preservation metadata may store technical information that supports preservation decisions and action, to document preservation action taken, to record the effects of preservation strategies, to ensure the authenticity of digital resources over time, and to note information about collection management and the management of rights. See the PADI website: <http://www.nla.gov.au/padi/topics/32.html>.

[83] Research Libraries Group (2002).

[84] This removal would be the functional equivalent of revoking the registration initially granted to the contribution on accession into the repository. In the journal publishing system, which integrates registration and certification, registration is most commonly denied by rejecting the paper for publication (that is, by denying certification).

[85] See, for example, <http://web.mit.edu/dspace/www/implementation/challenges.html>.

- Bass, Michael J. *et al*. 2002. DSpace: Internal Reference Specification: Technology and Architecture. Version 2002-03-01. Available from <http://dspace.org/technology/architecture.pdf>.
- Christopher Gutteridge. *EPrints 2.1 Documentation* (July 10, 2002). Available at: < http://software.eprints.org/documentation.php>.
- Preservation Metadata and the OAIS Information Model: A Metadata Framework to Support the Preservation of Digital Objects. The OCLC/RLG Working Group on Preservation Metadata. June 2002. See: <http://www.oclc.org/research/pmwg/>.
- Reference Model for an Open Archival Information System (OAIS). See: <http://www.ccsds.org/documents/pdf/CCSDS-650.0-B-1.pdf>.
- The Metadata Encoding and Transmission Standard (METS) schema is a standard for encoding descriptive, administrative, and structural metadata regarding objects within a digital library, expressed using the XML schema language of the World Wide Web Consortium. The standard is maintained in the Network Development and MARC Standards Office of the Library of Congress, and is being developed as an initiative of the Digital Library Federation. See: <http://www.loc.gov/standards/mets/>.
- The National Library of Australia's Preserving Access to Digital Information (PADI) initiative aims to provide mechanisms that will help to ensure that information in digital form is managed with appropriate consideration for preservation and future access. The PADI web site is a subject gateway to digital preservation resources. See: <http://www.nla.gov.au/padi/index.html>.
- *D-Lib Magazine* has published many articles that discuss elements of OAIS. The articles can be found by using the *D-Lib* search engine for the terms. See: <http://www.dlib.org/Architext/AT-dlib2query.html>.

Persistent Naming: The Handle System

For digital preservation, as well as for access and citation purposes, each object in the repository should have a unique and persistent reference identifier. Persistent identifiers, assigned to all material posted to the repository—and resolvable in perpetuity—would remain valid even were the repository content to be migrated to a new system or were management responsibility for the repository to be assigned to a third party.

Most institutional repositories will probably use the CNRI Handle System to achieve this continuity. The Handle System provides a comprehensive system for assigning, managing, and resolving persistent identifiers (known as "handles") for digital objects on the Internet. Handles can be used as Uniform Resource Names (URNs). Available at no cost, the Handle System includes an open set of protocols, a namespace, and an implementation of the protocols. The protocols enable a distributed computer system to store handles of digital resources and resolve those handles to locate and access the resources. The information associated with each handle can be changed to reflect the current state of the identified resource without changing the handle itself, thus allowing the name of the item to persist over changes of location and other state information. [86]

Several existing repository implementations assign persistent identifiers using the Handle System. DSpace uses the system to provide a storage- and location-independent

[86] On the CNRI Handle System, see: <http://www.handle.net/>. The principal mechanism for identifying, exchanging, and managing networked digital content amongst commercial publishers is the Digital Object Identifier (DOI) system. The DOI system, which itself uses the Handle system, provides a framework for managing intellectual content, linking content users with content providers, and enabling rights and copyright management for all types of digital media. Additionally, DOIs facilitate cross-reference document linking as implemented, for example, in the Open Citation Project and CrossRef. See < http://www.doi.org/>.

mechanism for creating and maintaining URLs. This model allows the repository to change its internal item retrieval mechanisms or physically move content without compromising reference citations and other links to the content.[87]

While the EPrints software automatically assigns a unique URL to each deposited object, these URLs would probably change if the content were migrated to another repository platform. At least one EPrints implementation has addressed this issue by using the CNRI Handle system to create a system that assigns a perpetual URL to each repository document.[88]

Interoperability & Open Access

For the repository to provide access to the broader research community, users outside the institution must be able to find and retrieve information from the repository. Therefore, systems must be able to support interoperability in order to provide access via multiple search engines and other discovery tools. An institution does not necessarily need to implement searching and indexing functionality to satisfy this demand: it could simply maintain and expose metadata, allowing other services to harvest and search the content. This simplicity lowers the barrier to repository operation for many institutions, as it only requires a file system to hold the content and the ability to create and share metadata with external systems.[89]

Interoperability requires persistent naming, standardized metadata formats, and a metadata harvesting protocol. Metadata describes the nature of the digital data stored in repositories (including the content, structure, and access rights administration). The metadata harvesting protocol allows third-party services to gather the metadata from distributed repositories and conduct searches against the assembled metadata to identify and ultimately retrieve documents. These mechanisms can be applied to any type of compliant digital library, creating a global network of digital research materials.[90]

The Open Archives movement spawned the Open Archives Initiative (OAI), which was established to develop and promote interoperability solutions to facilitate the dissemination of content.[91] The OAI is a collaborative effort to develop interoperability mechanisms that facilitate access to distributed digital content in the academic environment. The OAI provides the framework for facilitating the discovery of content in distributed repositories.

The OAI developed a set of interoperability standards called the OAI Protocol for Metadata Harvesting (OAI-PMH), which allows repositories to create metadata to describe content stored in the repository and make it available to others who wish to use

---

[87] See Bass *et al* (2002).

[88] Personal communication, Kim Douglas, Caltech Libraries, September 27, 2002.

[89] Personal communication, Herbert Van de Sompel, LANL, June 21, 2002.

[90] Detailed and specific metadata becomes increasingly expensive. To allow a lower level entry, the OAI supports a core set of metadata that represent a lowest common denominator. This lowers barriers to participation, and allows ephemera or other material that might not warrant the expense of extensive metadata tagging, while still adding value in terms of information retrieval. See Lagoze and Van de Sompel (2001) and Lynch (2001).

[91] "Open Archives" in this context requires some explanation. While many OAI proponents advocate monetarily free access to scholarly information, the OAI itself uses "open" to indicate machine interoperability, without a connotation of free or unlimited access. Additionally, for OAI, "archive" serves as a synonym for repository and does not necessarily indicate a digital preservation archive in the sense professional archivists might use the term. See: <http://www.openarchives.org/>.

it.[92] The OAI OAI-PMH supports the interoperability of digital repositories irrespective of type (institutional, discipline-specific, commercial, etc.) or content. The OAI maintains a list of OAI-compliant repositories from which OAI Service Providers can harvest metadata. To participate in this process, a repository must register with the OAI, once the institution's repository infrastructure is in place. The OAI certifies that a repository is fully compliant by validating the repository's metadata using a program that issues periodic OAI queries. Once these checks are complete, the OAI confirms the registration with the repository and adds the repository to the list of data providers.[93]

The OAI protocol requires that repositories offer the 15 metadata elements employed in unqualified Dublin Core metadata.[94] (See "Dublin Core Elements," in box below.) As a lowest common denominator, the unqualified Dublin Core will not be sufficiently detailed to serve the needs of many institutional repository collections.

However, the OAI protocol supports parallel metadata sets, allowing repositories to expose additional metadata specific to the repository's specific needs. Repositories that add domain-specific metadata sets to the Dublin Core should do so in consultation with other repositories to ensure a standardized presentation of these extended metadata sets.[95]

Dublin Core[96] simple or unqualified metadata includes:

- **Title**—the formal name of the resource;
- **Creator**—a person or corporate author of a resource;
- **Subject**—the topic of the resource, best expressed using a controlled vocabulary or other formal classification scheme;
- **Description**—an account of the resource's content—for example, an abstract or table of contents;
- **Publisher**—the entity responsible for making the resource available;
- **Contributor**—a person or corporate contributor to the resource's content;
- **Date**—Date that the resource was created, modified, or made available;
- **Type**—the nature or genre of the resource;
- **Format**—the physical or digital manifestation of the resource (for example, media type);
- **Identifier**—an unambiguous reference to the resource, best expressed in a manner conforming to a formal identification system (for example, DOI, URI, ISSN, ISBN, ISMN, etc.)
- **Source**—the resource from which the resource is derived;
- **Language**—the language of the resource;
- **Relation**—a reference to a related resource;
- **Coverage**—the geographical or temporal scope covered by the resource's content;
- **Rights**—information about rights held in and over the resource, including intellectual property rights, copyright, etc.

[92] A detailed description of the infrastructure can be found on the Open Archives Initiative web site. See also "Open Archives Initiatives Protocol for Harvesting Metadata," version 2.0, available at: <http://www.openarchives.org/OAI/openarchivesprotocol.html>.

[93] The EPrints software incorporates OAI configuration as part of the system configuration. See Gutteridge (2002), p. 16. The OAI web site provides a current list of data providers registered with the OAI. See <http://www.openarchives.org/Register/BrowseSites.pl>.

[94] See: <http://dublincore.org>. See also Lynch (2001).

[95] See <http://dublincore.org/documents/usageguide/>.

[96] For more on the Dublin Core elements, see: <http://dublincore.org/documents/dces/>.

Both the EPrints software and DSpace build metadata review and approval into their standard workflow processes, allowing contributors to specify baseline metadata based on the Dublin Core when submitting content, allowing the record to be checked and corrected prior to being made available publicly, and permitting users to search by this metadata. In addition to the baseline metadata, users can submit metadata specific to the item or to the collection of which it is a part (for example, metadata to indicate the relative location of images within a collection). However, this domain metadata may not be searchable by users.[97] EPrints also supports multilingual metadata.[98]

OAI-compliant Search Services

The OAI framework posits a publishing model that separates data providers (including institutional repositories) from service providers (metadata harvesters, search/retrieval, and other value-added access tools). Institutional repositories may serve both roles; however, they are considered logically discrete from an OAI perspective.[99] The full potential of institutional repositories and other digital archives requires the ability to federate these resources through a unified interface. Repository implementers should be aware of the federated search engines and other OAI-compliant service providers that, by harvesting the metadata of multiple repositories, will help leverage the value of each institution's repository content individually.

Several OAI-compliant search engines, mentioned below, are now available to supplement the local searching capability afforded by the EPrints and DSpace repository systems.

OAIster

The University of Michigan Libraries Digital Library Production Service launched version one of the OAIster (pronounced "oyster") search interface, in June of 2002. At the time of its release, OAIster was harvesting several hundred thousand records from over fifty institutions that made their records available via the OAI protocol.[100]

Arc

Arc, a federated searching service based on the OAI protocol, is a project of Digital Library Research group at Old Dominion University. Arc harvests metadata from several OAI compliant archives, normalizes them, and stores them in a search service based on a relational database (such as MySQL or Oracle). Although not yet a production service, Arc currently has hundreds of thousands of records, from various subject domains, from about 20 data providers.[101]

Citebase

Another OAI-compliant search service, under development as part of the Open Citation project and funded by the Joint NSF-JISC International Digital Libraries Research Programme,[102] is Citebase. In addition to harvesting metadata, Citebase harvests reference lists from large OAI archives, which it then uses to present citation-ranked search results. The search results can then be sorted by selectable criteria; such as how

---

[97] See Bass et al (2002) and Gutteridge (2002).

[98] Personal communication, Chris Gutteridge, University of Southampton, October 14, 2002.

[99] Shearer (2002). Available at: <http://www.carl-abrc.ca/projects/scholarly/open_archives.PDF>.

[100] See: <http://oaister.umdl.umich.edu/>.

[101] For more information, see the Arc web site (<http://arc.cs.odu.edu/>), and Liu *et al* (2001).

[102] For more on the Open Citation Project, see: <http://opcit.eprints.org/**>.**

many times a paper has been cited. While harvesting of OAI-compliant archives is currently limited, the project plans to cover more repositories moving forward.[103]

Interoperability & the Open Archives Initiative

Resources & Further Reading

- *D-Lib Magazine* has published many articles that discuss elements of OAI. The articles can be found by using the D-Lib search engine for the terms. See: <http://www.dlib.org/Architext/AT-dlib2query.html>.
- For more on the CNRI Handle System, see: <http://www.handle.net/>. The Corporation for National Research Initiatives (CNRI) is a not-for-profit organization that undertakes and promotes research centering around strategic development of network-based information technologies. See: <http://www.cnri.reston.va.us/>.
- Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) version 2 specification is available from: <http://www.openarchives.org/OAI/2.0/openarchivesprotocol.htm>.
- Kathleen Shearer. "The Open Archives Initiative: Developing an Interoperability Framework for Scholarly Publishing." CARL/ABRC Backgrounder Series #5 (March 2002). Available from <http://www.carl-abrc.ca/projects/scholarly/open_archives.PDF>.
  This paper provides a good overview of the inception and expansion of OAI (and allied initiatives) and its significance for the proliferation of interoperable digital repositories, as well as a description of the mechanisms that facilitate the interoperability of distributed repositories.
- A more detailed description of the OAI-PMH infrastructure can be found on the Open Archives Initiative web site (see: http://www.openarchives.org). See also "Open Archives Initiatives Protocol for Harvesting Metadata," version 2.0. Available from <http://www.openarchives.org/OAI/openarchivesprotocol.html>.
- Xiaoming Liu, Kurt Maly, Mohammad Zubair, and Michael L. Nelson. "*Arc* - An OAI Service Provider for Digital Library Federation." *D-Lib Magazine*, Volume 7, Issue 4 (April 2001). Available from <http://www.dlib.org/dlib/april01/liu/04liu.html>.

User Access & Rights Management

The repository's goal of long-term digital preservation does not necessarily mean that all content will be universally accessible in perpetuity. In addition to developing policies that define user communities, as discussed above, institutions must implement rights management systems that govern access to a repository's content.

Given the diverse formal and informal publishing practices amongst academic disciplines, an institution's content accession and access policies need to accommodate legitimate author concerns about access to pre- and post-publication material deposited in the repository. A variety of legitimate circumstances might require an institution to limit access to particular content to a specific set of users. These circumstances might include copyright restrictions, policies established by a particular research community (limiting access to departmental working papers to members of that department, for example), embargoes that an institution's Sponsored Programs Office might require to keep the institution in compliance with the terms of sponsor contracts, and even monetary access fees for certain data. Implementing these policy-based restrictions—which necessarily challenge to notion of "pure" open access for legitimate reasons—requires robust access and rights management mechanisms to allow or restrict access to content—and,

[103] For more information on Citebase, see: <http://citebase.eprints.org/>.

conceivably, to parts of digital objects—by a variety of criteria, including user type, institutional affiliation, user community, and others.[104]

Both the EPrints and DSpace systems allow any user to search and browse unrestricted repository content. However, the systems allow repository administrators—and their registered proxies, whether community-based or otherwise determined—the flexibility to control who can contribute, access, and update the digital resources posted to a repository. These access criteria can be based on a user's rights or community affiliation. Each system supports a user registration process and a secure process by which to administer user passwords. Additionally, DSpace intends to support commerce on subsets of a repository's contents.[105]

[104] The Shibboleth Project (see < http://middleware.internet2.edu/shibboleth/>) is addressing this cross-organizational sharing of web resources subject to access controls by developing architectures, policy structures, and practical technologies.

[105] See Bass *et al* (2002) and Gutteridge (2002).

SOURCES CITED

ALPSP. 1999. *What Authors Want: The ALPSP Research Study on the Motivations and Concerns of Contributors to Learned Journals*. (Worthing, West Sussex: The Association of Learned and Professional Society Publishers).

ALPSP. 2002. *Authors and Electronic Publishing: The ALPSP Research Study on Authors' and Readers' Views of Electronic Research Communication*. (Worthing, West Sussex: The Association of Learned and Professional Society Publishers).

Bass, Michael J. *et al*. 2002. DSpace: Internal Reference Specification: Technology and Architecture. Version 2002-03-01. Available from <http://dspace.org/technology/architecture.pdf>.

Bennett, Scott. 1999. "Authors' Rights." *Journal of Electronic Publishing* Volume 5, Issue 2 (December 1999). Available from < http://www.press.umich.edu/jep/05-02/bennett.html>.

Bentum, Maarten van. 2001a. "Authors' Attitudes and Perceptions and Strategies for Change with Respect to Electronic Publishing: A Literature Study." ARNO Report, Work Package 7 (March 2000). Available from <http://cf.uba.uva.nl/en/projects/arno/workpackages/arnowp7.rtf >.

Bentum, Maarten van. 2001b. "Attitude of Academic Staff and [Research] Managers to Electronic Publishing and the Use of Distributed Document Servers on University Level: A Survey Report." ARNO Report, Work Package 7 (November 2001). Available from <http://cf.uba.uva.nl/en/projects/arno/workpackages/arnowp7-survey.rtf>.

Bentum, Maarten van, Renze Brandsma, Thomas Place, and Hans Roes. 2001. "Reclaiming academic output through university archive servers." *New Review of Information Networking* (August). Available from <http://cwis.kub.nl/~dbi/users/roes/articles/arno_art.htm>

Bjork, Bo-Christer, and Ziga Turk. 2000. "How Scientists Retrieve Publications: An Empirical Study of How the Internet Is Overtaking Paper Media." *The Journal of Electronic Publishing* Volume 6, Issue 2 (December, 2000). Available from <http://www.press.umich.edu/jep/06-02/bjork.html>.

Crow, Raym. 2002. The Case for Institutional Repositories: A SPARC Position Paper. (Washington, DC: Scholarly Publishing & Academic Resources Coalition). Available from <http://www.arl.org/sparc/IR/IR_Final_Release_102.pdf>.

Ginsparg Paul. 2001. "Creating a Global Knowledge Network." Conference on Electronic Publishing in Science, Paris (February 20, 2001). Available from <http://arXiv.org/blurb/pg01unesco.html>.

Gutteridge, Christopher. 2002. *EPrints 2.1 Documentation* (July 10, 2002). Available from <http://software.eprints.org/documentation.php>.

Kling, Rob, and Roberta Lamb. 1996. "Analyzing Alternate Visions of Electronic Publishing and Digital Libraries." In *Scholarly Publishing: The Electronic Frontier*. Edited by Robin P. Peek and Gregory B. Newby (Cambridge, Mass.: MIT Press): 17-54.

Kling, Rob, and Geoffrey McKim. 2000. "Not Just a Matter of Time: Field Differences and the Shaping of Electronic Media in Supporting Scientific Communication." *Journal of the American Society for Information Science*. Volume 51, Number 14: 1306-1320.

Kling, Rob, Lisa Spector, and Geoffrey McKim. 2002. "Locally Controlled Scholarly Publishing via the Internet: The Guild Model." CSI Working Paper no. WP-02-01 (June 2002). Available from < http://www.press.umich.edu/jep/08-01/kling.html>.

Lagoze, Carl, and Herbert Van de Sompel. 2001. "The Open Archives Initiative: Building a low-barrier interoperability framework." *Joint Conference on Digital Libraries 2001*. Available from <http://www.openarchives.org/documents/oai.pdf>.

Lavoie, Bruce. 2000. "Meeting the Challenges of Digital Preservation: The OAIS Reference Model." *OCLC Newsletter*, No. 243: pp. 26-30.

Lawrence, Steve. 2001. "Online or invisible?" *Nature* 411 (6837): 521.

Liu, Xiaoming, Kurt Maly, Mohammad Zubair, and Michael L. Nelson. 2001. "*Arc*—An OAI Service Provider for Digital Library Federation." *D-Lib Magazine* Volume 7, Issue 4 (April 2001). Available from <http://www.dlib.org/dlib/april01/liu/04liu.html>.

Lynch, Clifford A. 2001. " Metadata harvesting and the Open Archives Initiative." *ARL* 217 (August).

McMillan, Gail, Edward A. Fox, and John L. Eaton. 1999. "The Evolving Genre of Electronic Theses and Dissertations." 1999 Hawaii International Conference on System Sciences.

Nixon, William J. 2002. "The evolution of an institutional e-prints archive at the University of Glasgow." *Ariadne* 32 (July 8, 2002). Available from <http://www.ariadne.ac.uk/issue32/eprint-archives/>.

Pinfield, Stephen. 2001. "How Do Physicists Use an E-Print Archive?" *D-Lib Magazine* Volume 7, Issue 12. Available from <http://www.dlib.org/dlib/december01/pinfield/12pinfield.html>.

Pinfield, Stephen, Mike Gardner, and John MacColl. 2002. "Setting up an institutional e-print archive" *Ariadne* 31. Available from <http://www.ariadne.ac.uk /issue31/eprint-archives/intro.html>.

Research Library Group. 2002. *Trusted Digital Repositories: Attributes and Responsibilities*. An RLG-OCLC Report. (Mountain View, CA: Research Libraries Group).

Shearer, Kathleen. 2002. "The Open Archives Initiative: Developing an Interoperability Framework for Scholarly Publishing**."** CARL/ABRC Backgrounder Series #5 (March 2002). Available from <http://www.carl-abrc.ca/projects/scholarly/open_archives.PDF>.

Tenopir, Carol and Donald W. King. 2000. *Towards Electronic Journals* (Washington, DC: SLA Publishing).

Weintraub, Irwin. nd. "The Role of Grey Literature in the Sciences." Available at: <http://library.brooklyn.cuny.edu/access/greyliter.htm>.

APPENDIX: SELECT LIST OF INSTITUTIONAL REPOSITORIES

A growing number of institutions and consortia are actively engaged in setting up and running institutional repositories. The practical experiences gained by these initiatives—organizational, technical, and legal—should prove instructive to other institutions.

The select list below includes repositories that are institutional in scope and that contain multiple document types. Thus, it excludes discipline-specific e-print servers and university repositories that contain only theses and dissertations. Lists that include those types of repositories can be found elsewhere.[106]

AUSTRALIA

Australia National University
E-Print Repository
Content: preprints, published articles, theses and dissertations, etc.
System software: Eprints.org
URL: <http://eprints.anu.edu.au/>

CANADA

Université de Montréal
Papyrus—Institutional Eprints Archive
Content: preprints, published articles
System software: Eprints.org
URL: <http://papyrus.bib.umontreal.ca/>

DENMARK

Aalborg University
Electronic Library
Content: preprints, published articles; PDF only
System software: Unknown
<http://www.aub.auc.dk/phd/mainpage.html>

FRANCE

Institut Jean Nicod
Archive Electronique
Content: preprints, published articles (in journals and anthologies), published correspondence.
System software: Eprints.org
URL: <http://jeannicod.ccsd.cnrs.fr/>

GERMANY

Universität Dortmund
Eldorado
(in German)

[106] See, for example, <http://www.signal-hill.org/nav/archives2.html> and <http://software.eprints.org/#sites>.

Content: preprints, published articles (in journals and anthologies), published correspondence, etc.
System software: Hyperwave (<http://www.hyperwave.com/e/>)
URL: <http://eldorado.uni-dortmund.de:8080/rootcollection;internal&action=buildframes.action>

Universität Essen
MILESS
(in German)
Content: preprints, published articles, teaching materials, theses and dissertations, etc.
System software: MyCoRe (<http://www.mycore.de/projektbeschreibung.html>)
URL: <http://miless.uni-essen.de/>

Universität Stuttgart
OPUS (Online Publications University of Stuttgart)
Content: preprints, published articles, teaching materials, theses and dissertations, etc.
System software: OPUS System[107]
URL: <http://elib.uni-stuttgart.de/opus/doku/english/index.html>

Universität Konstanz
KOPS-Datenbank Konstanzer Online-Publikations-System
Content: preprints, published articles, teaching materials, theses and dissertations, etc.
System software: OPUS System
URL: <http://www.ub.uni-konstanz.de/kops/>

India

Indian Institute of Science (Bangalore)
eprints@iisc
Content: preprints, published articles
System software: Eprints.org
URL: <http://eprints.iisc.ernet.in/>

Italy

Università degli studi di Firenze
E-prints archive
Content: preprints, published articles
System software: Eprints.org
URL: <http://biblio.unifi.it/indexeng.html>

The Netherlands

University of Maastricht
E-prints archive
Content: primarily research papers
System software: Eprints.org
URL: < http://137.120.22.236/www-edocs/default.asp?taal=ENG&webnaam=edocs >

[107] Other institutions said to be implementing the OPUS system include: Universities of Brunswick, Freiburg, Heidelberg, Hohenheim, Mannheim, Regensburg, Saarbruecken, Tuebingen. OPUS is currently being tested by the Universities of Bamberg, Bayreuth, Gießen, Goettingen, and Passau.

Utrecht University
Dispute
Content: small but fully operational repository, containing a subset of all Utrecht publications (approximately 800 full text articles) and the collection of Utrecht online dissertations (approximately 300 dissertations).
System software: custom (?)
URL: <http://dispute.library.uu.nl/>

Sweden

Blekinge Institute of Technology
Electronic Research Archive
Content: currently research papers
System software: custom (?); PDF format
URL: <http://www.hk-r.se/fou/>

Lulea Institute of Technology
Publications from LTU
Content: research papers, theses, and dissertations
System software: custom (?)
URL: <http://epubl.luth.se/index-en.html>

Lunds Universitet
Lunds University Library Full-Text Project (LUFT)
Content: teaching material, report series, and research papers
System software: custom
URL: <http://www.lub.lu.se/luft/>

Switzerland

CERN Scientific Information Service
CERN Document Server (CDS)
Content: preprints, research papers, books, photographs, video clips, etc.
System software: custom
URL: <http://cds.cern.ch/>

U.K. & IRELAND

National University of Ireland, Maynooth
NUI Maynooth Eprint Archive
Content: preprints, research papers
System software: Eprints.org v. 2.0
URL: <http://eprints.may.ie/>

University of Bath
ePrints@Bath
Content: preprints, published articles, theses and dissertations, etc.
System software: Eprints.org
URL: <http://eprints.bath.ac.uk/>

University of Glasgow
EPrints at Glasgow
Content: preprints, published articles, theses and dissertations, etc.
System software: Eprints.org
URL: <http://eprints.lib.gla.ac.uk/>

University of Nottingham
Nottingham ePrints
Content: preprints, published articles, theses and dissertations, etc.
System software: Eprints.org
URL: <http://www-db.library.nottingham.ac.uk/eprints/>

University of Strathclyde
StrathPrints
Content: preprints, published articles, theses and dissertations, etc.
System software: Eprints.org
URL: < http://eprints.cdlr.strath.ac.uk/>

USA

California Digital Library
eScholarship
Content: preprints, published articles, theses and dissertations, etc.
System software: custom with Berkeley Electronic Press (bepress)
URL: <http://escholarship.cdlib.org/wprepositories.html>

Caltech
CODA: Caltech Collection of Open Digital Archives
Content: preprints, published articles, theses and dissertations, etc.
System software: Eprints.org
URL: <http://coda.caltech.edu>

MIT
DSpace
Content: preprints, published articles, theses and dissertations, etc.
System software: DSpace infrastructure software
URL: < https://hpds1.mit.edu/index.jsp>

Hofstra University
Hofprints—Hofstra University E-Print Archive
Content: preprints, published articles, theses and dissertations, etc.
System software: Eprints.org
URL: <http://hofprints.hofstra.edu/>

Virginia Tech, Digital Library and Archives
Digital Library & Archives
Content: preprints, published articles, theses and dissertations, etc.
System software: custom (?)
URL: <http://scholar.lib.vt.edu/DLASPS/index.html>