

UC San Diego

UC San Diego Previously Published Works

Title

In Search Of A Standardized Model for Institutional Repository Assessment or How Can We Compare Institutional Repositories?

Permalink

<https://escholarship.org/uc/item/9w1910rv>

Authors

Thomas, Chuck
McDonald, Robert H.

Publication Date

2008-11-01

Peer reviewed

In Search Of A Standardized Model for Institutional Repository Assessment
or
How Can We Compare Institutional Repositories?

I. Introduction

"An institutional repository concentrates the intellectual product created by a university's researchers, making it easier to demonstrate its scientific, social and financial value. Thus, institutional repositories complement existing metrics for gauging productivity and prestige...this demonstration of value can translate into tangible benefits, including the funding...that derives in part from an institution's status and reputation." (Crow, 2002, p. 6)

Assessing universities and faculty is a continuous struggle. Academic administrators must labor year after year to gather meaningful statistics for assessment exercises such as periodic institutional accreditations, program reviews, and annual funding requests. It is hard to overstate the difficulty and complexity of compiling such data. The professional literature of higher education administration contains frequent calls over the past several decades, for better ways to measure performance in colleges and universities.

Despite this recognized need, few tools or standards have emerged to fill the void. The U.S. Department of Education's *Report on the Future of Higher Education in the United States* (2006, p. 4) noted, "We have found a remarkable shortage of clear, accessible information about crucial aspects of American colleges and universities...this lack of useful data and accountability hinders policymakers and the public...and prevents higher education from demonstrating its contribution to the public good." Norris, *et al.* (2008, p. 44) note that the lack of performance data for U.S. colleges and universities is particularly problematic today, when public demand for such data is escalating.

In recent years, some Institutional Repository (IR) advocates have emphasized the potential utility of IRs in institutional assessment. Their potential to provide "online,

continuous, metrics-based" (Harnad, 2006) scholarly performance and impact measurements is a persuasive argument for implementing IRs. On-demand metrics would be an important asset for academic administrators and faculty at all levels within a university. An institutional repository, if properly planned, deployed and supported via appropriate policies and resources, could provide real-time reports demonstrating productivity, impact and overall value for research organizations.

Achieving the full administrative benefit of IRs, however, is hindered today by ongoing evolutions within the realms of digital repositories, individual and disciplinary scholarly communication behaviors, and policies of organizations that support research. Before IRs can serve as tools for institutional and faculty assessment, universities must reach a greater degree of agreement about the purpose, content and shared characteristics of IRs. Currently, ideological, social, financial, and technical shifts underway make it difficult to reach consensus about what should be measured, by whom, and for what purpose.

Additionally, as Thomas & McDonald (2007) observed, digital repositories and other scholarly communication tools are growing more hybridized and more interconnected. If this trend continues, identifying the scope, content and context of individual repositories will be an increasingly imprecise exercise. Though institutional, disciplinary and other scholarly digital repositories certainly will be important parts of tomorrow's scholarly communication fabric; today they differ so much from each other that the statistics and measures they can produce do not conform to any common standards or definitions. In this environment, measurements and statistics produced by IRs are largely meaningless for comparing individuals, groups or organizations across universities.

One of the many steps needed toward better comparison of IRs, is a shared understanding of how to evaluate them. IRs are more than just particular software, policies, or content. Instead, they are enterprise-wide programs. Measuring an IR's progress and impact might be the best way to build faculty support for an existing local repository, if only meaningful and standardized evaluative metrics were available (Westell, 2006). Proudman (2008) reminded us a standardized evaluative framework also would be valuable for encouraging more institutions to implement IRs.

This paper summarizes some of the existing candidate frameworks for IR

assessment from a futurist's perspective, and considers potential challenges on the path to establishing comparative assessment metrics. The authors consider what a future IR analysis might look like, and how it would differ from the evaluative frameworks being discussed today.

II. Existing Candidate Frameworks for IR Evaluation

Currently, the task of evaluating digital scholarly repositories is complicated by variations in:

- what repositories contain;
- who funds and administers each;
- underlying legal, social and policy infrastructure for each repository;
- who contributes to the repository; and
- motivations for contributing, whether they be mandates, disciplinary cultural norms, or other incentives

Additionally, distinctions made in relevant literature between *institutional*, *disciplinary*, and other types of repositories, can quickly become murky when one surveys the global network of databases and systems that hold content and metadata, sometimes contain only citations and links to content and metadata in other repositories, often overlap significantly with other sites, and are part of a scholarly communication web where individual scholars in different disciplines make choices on an item-by-item basis to deposit their works in any one of multiple possible repositories.

With these and other complications, it is difficult to set reliable guidelines or tests for classifying repositories as institutional, disciplinary, or other types. Instead of using rigid one-dimensional groupings, it is likely in coming years that repositories will be described according to their mapped plots on several multi-dimensional continua. These continua might include dimensions such as ownership and management, communities served directly and indirectly by the repository, level of usage by both contributors and by researchers, scope of content in the repository, underlying technologies and policies, and relationships and interoperation with other repositories. Even if the *institutional* and *disciplinary* categories do not prove useful over the long-term, they remain

useful distinctions today for those considering scholarly repository evaluative criteria.

The easiest example of differences between disciplinary and institutional repositories is to consider the hypothetical search and discovery experiences of someone looking for scholarly information in each. Users of a functioning recognized domain or disciplinary repository could reasonably expect to discover full-text reports, citations, or other references to most of the current research within that specific field. In many ways, this search and discovery process would equal what could be expected from a modern research library that collects comprehensively within the same content domain. In both the disciplinary repository and the research library, of course, the completeness and comprehensiveness of what could be discovered depends on how successful each is in its acquisition routines.

Now, contrast that experience with what the same user would encounter when instead searching a disciplinary repository. Imagine searching a campus library, and discovering the library provides access to only materials produced by campus faculty and departments, who may or may not have incentives to share their work with the library, due to its very local and limited influence on their wider academic disciplines. Researchers trying to conduct research or understand current work within any field would have a very difficult time if they had to rely solely on what they could find in such a library.

This example is obviously very simplistic. Deposit in a disciplinary or institutional repository does not preclude simultaneous deposit in other repositories, publication in journals, and dissemination through other scholarly communication channels as well. As the web of repositories becomes more complex, scholars who deposit works in one repository may one day be able to expect that metadata-sharing and other repository services will automatically push their works to the attention of other external channels. However, the main point to understand from the example above is that institutional and disciplinary repositories exist for fundamentally different purposes, and therefore require different evaluative criteria. Though advocates of open access to research legitimately champion both types of repositories as disruptors of existing barriers,

Though frameworks do exist for some specific aspects of digital repositories, including the OCLC/NARA/CRL (2007) *Trustworthy Repositories Audit and Certification* checklist for preservation repositories, and the Open Society Institute's (2004) technical comparison of

repository softwares, no general evaluative models have emerged for disciplinary or topic-based scholarly repositories.

For institutional repositories, however, three analytical/comparative frameworks recently surfaced in the scholarly literature. Proudman's (2008) analysis of European institutional repositories employed the following evaluative criteria, identified as "recurring themes of international and national discourse on the issue of open access and scholarly communication":

- Policies;
- Organization;
- Mechanisms and influences for populating repositories;
- Services;
- Advocacy & communication;
- Legal issues

Proudman applied this evaluative framework to multiple institutional repositories, and concluded these categories cannot provide a numeric score, but instead are useful to compile a qualitative profile of a repository's strengths and weaknesses. No individual evaluative category was found to be most indicative of a successful repository.

Like Proudman, Westell (2006) had earlier applied a framework of several evaluative criteria, and found most of them tend to generate qualitative, not quantitative, success indicators. Her study used eight categories, including:

- Repository mandate;
- Integration with institutional planning;
- Funding model;
- Relationship with digitization centers;
- Interoperation;
- Content measurement;
- Promotion;
- Preservation strategy

Interestingly, Westell deemed "user acceptance" of a repository as another potential category, but one worthy of its own separate study. In contrast to Proudman's subsequent findings, Westell found one indicator, the amount of content in a repository, to be the

greatest determinant of whether a repository is considered successful or not (p. 223).

Both of these studies revealed the problems inherent in applying a set of common evaluative categories to a heterogeneous group of repositories. Many of the analytical categories used in each study are not conducive to consistent and accurate measurement across repositories. In their own ways, however, each study illustrated the potential and the utility of measuring success of IRs.

Recently, Kim & Kim (2007) reported on their efforts to develop an evaluative framework for IRs in South Korea's university system. By synthesizing literature on evaluation of both IRs and digital libraries, through analysis of six well-known repositories, and through extensive testing and interviews with IR experts, they developed a rubric of four broad evaluative categories and numerous indicators within each category. This analytical framework was tested on a single digital repository, but evidently soon will be tested on a wider group of IRs. In its present manifestation, this evaluative framework is broken into the following categories:

- Content (Diversity, Currency, Size, Metadata)
- System and network (Interoperability, Use of help services like FAQ and Q&A)
- Use, users and submitters (Use ratio, User satisfaction, Submitter satisfaction, User/Submitter support)
- Management and policy (Budget, Staffing, Library awareness of Open Access and related issues, Copyright management, IR Marketing, Institutional support, Policies and procedures in place, Diversity of archiving methods)

By the authors' own admission, some of these analytical criteria cannot be applied consistently across multiple repositories. Also, though the Kim & Kim framework provides some quantitative measurements, much of the data it provides is not the quantitative sort that encourages easy comparison of repositories. Additionally, their framework makes multiple assumptions about the administrative and organization structure of a university and its IR, and about the role of libraries and librarians as indicators of a successful IR effort. From a librarian's perspective, these assumptions are an ego booster, but the Kim & Kim model will likely require significant modifications to be applicable for use in U.S. and North American environments.

III. The Need for Measurement and Comparison

Each of the three studies described above produced useful descriptions of individual repositories, such as case studies of each repository analyzed, and lists of unique or interesting characteristics of IR. This kind of detailed contextual narrative is valuable, and in the Westell and Proudman studies, also resulted in lists of useful tips, best practices and recipes for better repositories. However, none of the investigations discussed earlier claim to be scoring systems for measuring and comparing repositories. Without the ability to compare, repository administrators may have a difficult time demonstrating the significance or degree of an IR's success.

Kyrillidou (2008, p. 11) reminds us qualitative, more subjective descriptions of digital library collections may soon replace the more easily-comparable quantitative measures (e.g., total budget expenditures and item counts) used in the past. Nonetheless, hard numbers and comparisons will continue to be strongly preferred or required to secure ongoing administrative and financial support for most repositories. One strategy for achieving effective evaluative frameworks for IRs, is to adopt many of the qualitative criteria explored by Westell, Proudman, and Kim & Kim, but to supplement them with repository-wide quantitative measurements borrowed from other statistics used by scholars and university administrators. Some of these might include:

- Scholarly impact of both individual digital documents and the repository overall;
- Comparisons of resource Inputs vs Outputs;
- Categorized totals amount of content (e.g., published research gray literature);
- Correlated measures of productivity (e.g., # of faculty, # of deposits per scholar);
- Relationship and influence of local IR with disciplinary repositories, journals, etc.;
- Indicators and adjustments for overall organizational size and resources.

These are just some examples of the numbers that would be useful for administrative evaluation of an IR. Librarians and others involved with managing IRs could benefit by becoming better acquainted with other evaluative frameworks used in upper-level academic administration, such as for regional and disciplinary academic accreditations, and for higher education reporting to the U.S. Dept. of Education (Cocklin, 2008). All of these evaluation exercises, which university administrators must manage on a continuing,

cyclical basis, are based on very general categories similar to the IR evaluation frameworks described earlier. However, university administrators have learned how to operationally integrate definitions and break evaluative criteria into quantitative measures. In the same way, IR managers must recognize the need for reports and statistics that help measure and compare success of IRs.

IV. Future Evolution of Institutional Repositories and Evaluation

The landscape of digital scholarly repositories is still evolving, and no one can say for certain how long the current categories of "institutional" or "disciplinary" repositories will persist. Much of the uncertainty about the future is due to an ongoing decentralization of power and authority in scholarly communication. Institutions, publishers, and scholarly societies that once exercised extensive control over scholarly communication are losing their grip as grassroots; scholar-driven initiatives are gaining momentum. The SCOAP3 initiative within the high energy physics community is a perfect example of such developments.

As individuals and like-minded clusters of researchers exploit new technologies and redefine scholarly communication, new types of digital scholarly repositories may emerge and disrupt existing classifications. Calabrese (1987, p. 11-16) described "problem-centered" multidisciplinary groups that are defined only by the research questions they investigate; might we see problem-centered digital depositories emerge as the Next Big Thing in this field? If so, will they have their own particular evaluative needs?

Whatever changes may occur, it is important to keep several important points in mind when considering evaluation of institutional repositories:

1. Though research questions may transcend particular scholars or organizations, individual researchers are still usually paid and supported by particular institutions.
2. Continuing support of scholars and programs is dependent on continuing demonstrations of performance and impact.
3. The changing scholarly communication landscape gives scholars many choices on where they may publish or share their intellectual output.
4. Scholars are still averse to any extra work of depositing their work in prescribed repositories (Sale, 2007).
5. The tasks of counting and measuring performance and impact, while extremely important to institutions and funders, are arguably now more difficult thanks to points 3 and 4.

If institutionally-hosted digital repositories do not survive as a viable component of scholarly communication, this would not be surprising. Their primary utility is for the benefit of the institutions and not as a primary node where researchers would be expected to conduct comprehensive literature searches. Regardless of where scholars' works reside, however, institutions and funders will want to count, measure and compare the quality and quantity of research they sponsor. At some point, then, the enormous potential utility of an IR, or for automated tools that track and count scholarly deposits in diverse distributed repositories, will continue to resurge, and probably will lead to discussions of institutional mandates for secondary deposits by locally-supported scholars in locally-supported repositories.

At this point, we come back full-circle to the question of how to evaluate locally-supported repositories, but now with a better understanding of their context. Earlier, this paper discussed the possible tension between institutional administrators' needs and desires for quantitative measurements, and the emerging consensus that qualitative descriptions are needed to fully understand and gauge an institutional repository's success. By combining some of both, perhaps IRs can be adequately measured and compared against each other.

However, in the search for new metrics in the new realms of digital collections, it might also be constructive to consider that comparisons do not always have to mean "more vs. less" or "better vs. worse". Instead, if we remember that each IR is a complex, unique

combination of local policies, resources and specializations, intentions and needs, then each IR may be more like a particular color value on a palette of thousands or even millions of possible colors, where each color value is the result of a unique combination of light from the red, green and blue spectrums. For purposes of comparison, then, one possible alternative to a multi-dimensional graphical plot of a repository's attributes, might be as simple as a simple color palette, where one can produce a unique color value associated with an IR by inputting its value or ranking in various spectra, in the same way that the user of image software might choose a color through a simple interface like the one shown below in Figures 1-3.

V. Conclusion

The future of IRs is uncertain, but the need for better institutional assessment tools is undisputed. If IRs survive the ongoing evolutions within scholarly communication, it will undoubtedly be because of their utility as tools for measuring and comparing faculty and organizational performance. In turn, IRs will also need to be evaluated and compared as programmatic activities within universities and colleges. The evaluative criteria of IRs are different from other types of digital scholarly repositories, because they serve a different purpose.

Librarians and academic administrators should remember that scholars are not waiting on them to make decisions. Instead they are re-inventing the systems of scholarly communication, and as often as not, not including outsiders in their decisions. Simultaneously, lines between institutional, disciplinary, and other digital scholarly repositories continue to change. The net result is a lag between the current scholarly communication landscape, and the discourse (published literature, proposed frameworks, etc.) on how to evaluate components such as IRs.

Recent efforts to develop IR evaluative frameworks have produced criteria that favor qualitative assessments, due to the variety of policies, resources, organizations and scholars unique to each institution. This creates a tension because traditional metrics and evaluative criteria for both libraries and higher education have focused on quantitative measurements. As libraries, university administrators, and IR managers seek ways to

evaluate the success of IRs, they should blend together both quantitative and qualitative measurements, and devise innovative representations that assign precise, accurate and rich indicators of IR attributes.

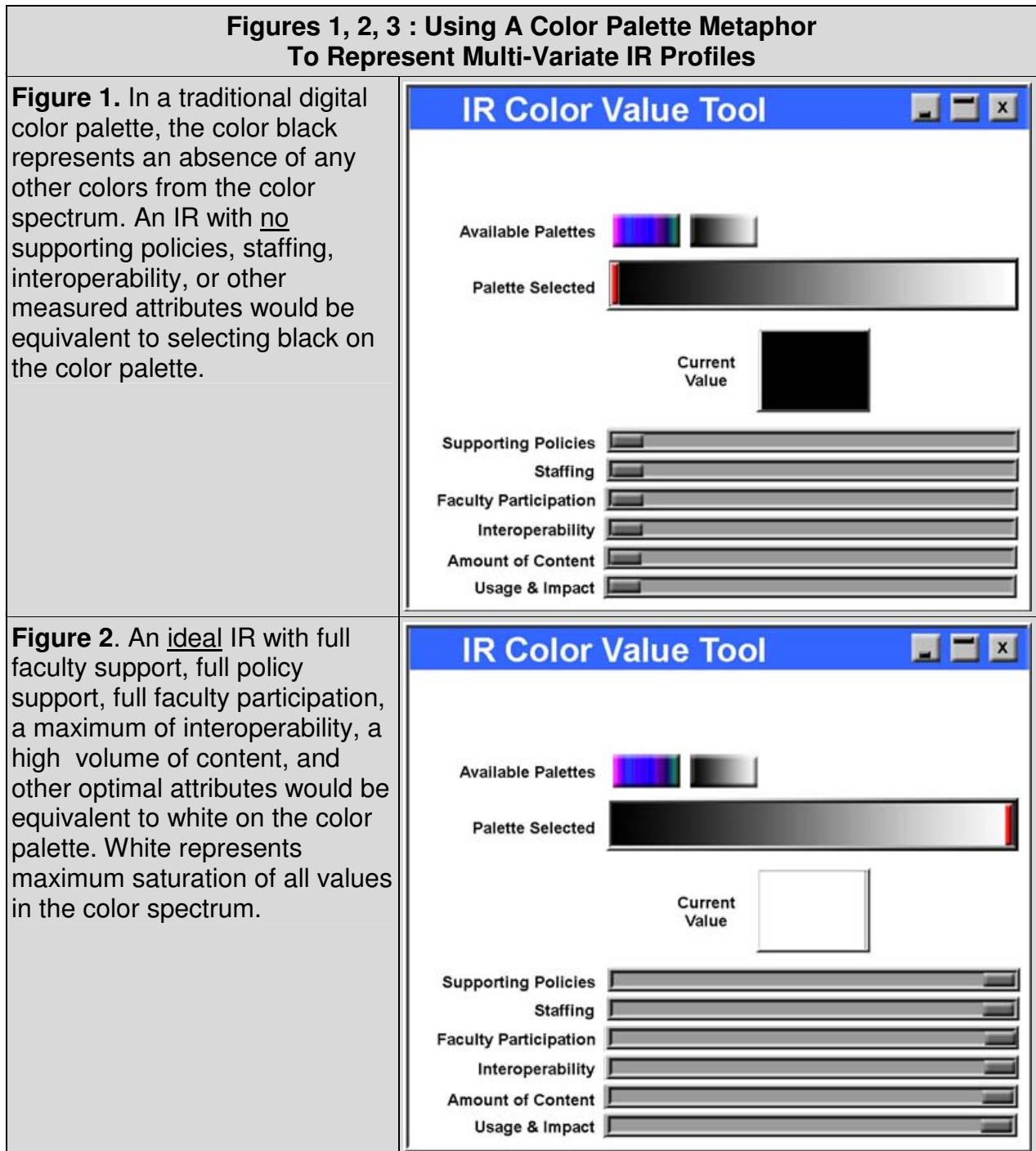
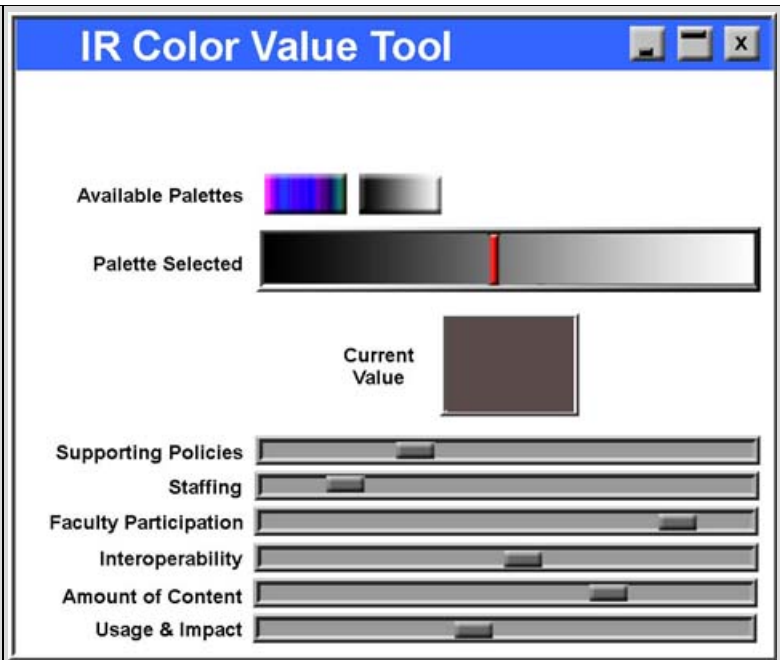


Figure 3. In this mockup of a color-value calculation tool, users could select a palette of millions of colors with more variable attributes to input, or a grayscale palette with a relatively smaller set of thousands of possible colors. A typical IR might score low in some attributes, and higher in others, to end up with a particular shade of gray as its corresponding aggregate color-value assignment.



V. Sources Cited

Calabrese, A. M. (1987). Social science discourse: Issues in scholarly communication. (Paper presented at the 73rd Annual Meeting of the Speech Communication Association, Boston, Nov. 5-8).

Center for Research Libraries. (2007-2008). *Trustworthy Repositories Audit and Certification*. <<http://www.crl.edu/content.asp?l1=13&l2=58&l3=162&l4=91>>.

Cocklin, J. (2008). *Adding Context to Academic Library Assessment: Using the Integrated Postsecondary Education Data System for Institutional and Comparative Statistics*. (Presentation at ARL Library Assessment Conference, Seattle, Aug. 4-7).

Crow, R. (2002). *The case for institutional repositories: A SPARC position paper*. Washington, DC: The Scholarly Publishing and Academic Resources Coalition. <http://www.arl.org/sparc/bm~doc/ir_final_release_102.pdf>.

Harnad, S. (2006). Online, continuous, metrics-based research assessment. (Unpublished technical report). <<http://eprints.ecs.soton.ac.uk/12130/1/rae-metric.html>>.

Kim, H. H. and Kim, Y. H. (2007). An evaluation model for the national consortium of institutional repositories of Korean universities. In *Proceedings of the American Society for Information Science and Technology*. (London: John Wiley & Sons), <<http://dx.doi.org/10.1002/meet.1450430176>>.

Kyrillidou, M. (2008). Reshaping ARL statistics to capture the new environment. *ARL Newsletter* 256. <<http://www.arl.org/bm~doc/arl-br-256-stats.pdf>>.

Norris, D., Baer, L., Leonard, J., Pugliese, L., and Lefrer, P. (2008). Action analytics: Measuring and comparing performance that matters in higher education. *EDUCAUSE Review* Jan./Feb. 2008, p. 43-67.

Open Society Institute. (2004). A guide to institutional repository software (3rd ed.) <http://www.soros.org/openaccess/pdf/OSI_Guide_to_IR_Software_v3.pdf>.

Proudman, V. (2008). The population of repositories. In Eds. K. Weenink, L. Waaijers and K. van Godtsenhoven, *A DRIVER's Guide to European Repositories*. (pp. 49-101) Amsterdam: Amsterdam University Press.

Sale, A. (2007). The patchwork mandate." *D-Lib Magazine* 13:1/2. <[doi:10.1045/january2007-sale](http://dx.doi.org/10.1045/january2007-sale)>.

Sponsoring Consortium for Open Access Publishing in Particle Physics. (2008). Toward open access publishing in high energy physics [website]. <<http://scoap3.org/>>.

Thomas, C., and McDonald, R. (2007). Measuring and comparing participation patterns in digital repositories. *D-Lib Magazine* 13:9/10. <<http://repositories.cdlib.org/postprints/2540>>.

U.S. Dept. of Education (2006). *A test of leadership: Charting the future of U.S. higher education* (Pre-publication copy, Sept.).

<<http://www.ed.gov/about/bdscomm/list/hiedfuture/reports/pre-pub-report.pdf>>.

Westell, M. (2006). Institutional repositories: Proposed indicators of success. *Library Hi Tech*, 24(2), 211-226.