Measuring Value in Open Access Repositories

MEGAN WACHA and MEREDITH WISNER

School of Information and Library Science, Pratt Institute, New York, New York, USA

Keywords

Open Access, Institutional Repositories, Scholarly Communication, Academic Libraries, Electronic

Resources, Digital Libraries

Abstract

Open access institutional repositories were created to promote access to information, encourage scholarly

communication, and demonstrate institutional prestige. While these repositories have been widely

adopted, the quality of their contents often fails to represent their institution's scholarly

output. Moreover, current research uses measurements of quantity, not quality, to assess their value. In

response, this article opens new areas of scholarly inquiry by assessing the quality of contents. This is

accomplished through a cross-sectional study of repositories at American colleges and universities across

the academic spectrum, using citation indexing to identify an institution's articles and authors of highest

impact.

Introduction

As defined by the Scholarly Publishing & Academic Resources Coalition position paper from 2002, open

access institutional repositories (OARs) were created to encourage scholarly communication outside

traditional publishing models, demonstrate the prestige of institutions by highlighting their scholarly

output, and to make this output accessible to the wider academic community.[1] However, while

institutional repositories have been adopted across the academic spectrum, the quality of the materials maintained within them is not often representative of the institution's academic stature. In order to increase access to quality materials and create real alternatives to journal publication, open access repositories must contain materials of value, both to serve needs of academic institution as well as the larger scholarly community. This article will measure the value of institutional repositories through an evaluation of their content as representative of the institution's intellectual output.

Literature Review

To frame our conception of institutional repositories, we looked to Clifford Lynch, Director of the Coalition for Network Information, who describes them in his 2003 article *Institutional Repositories: Essential infrastructure for Scholarship in the Digital Age:*

> ... a university-based institutional repository is a set of services that a university offers to the members of its community for the management and dissemination of digital materials created by the institution and its community members. It is most essentially an organizational commitment to the stewardship of these digital materials, including long-term preservation where appropriate, as well as organization and access or distribution.[2]

We then consulted a few important investigations into the creation and implementation of institutional repositories in academe. The first rigorous treatment of the topic, conducted in 2005 by Lynch and Joan Lippincott, provides statistics on the number of institutional repositories in doctoral universities in the United States, as well as the kinds of materials they actively collect.[3] Cat S. McDowell builds upon this seminal article in 2007; covering all levels of academic institutions as defined by the Carnegie Classification.[4] Other studies include The Association of Research Libraries' *SPEC Kit* on institutional repositories from 2006[5] and the *Census of Institutional Repositories in the United States: MIRACLE Project Research Findings* by the Council on Library and Information Resources from 2007.[6] The latter includes a national census study into the use of institutional repositories in the U.S., library staff interviews, detailed investigations into five institutional repositories, a user study and a study investigating information retrieval.

Because we were specifically interested in measuring the value of open access repositories, we also consulted articles dealing solely with OARs. These articles tended to be narrowly focused; including specific disciplines using OARs, like Mohammad Hanief Bhat's 2009 article on OARs in computer science;[7] the utilization of OARs in countries outside the United States, such as Bhat's, work in India (2009),[8] Rowan Brownlee's in Australia (2009),[9] Stevan Harnad's in the United Kingdom, (2005)[10] and Blanca Rodriguez Bravo and Maria Luisa Alvite Díez's work in Spain (2007);[11] as well as the quality and ability to harvest OAR metadata, as with the work of John Chapman, David Reynolds and Sarah Shreeves (2009),[12] Jessica Branco Colati, Robin Dean, and Keith Maull (2009),[13] and Jung-Ran Park (2009).[14] In many cases, the above resources focused on the quantity of work deposited, but paid less attention to the quality and type of materials being collected.

Bhat's 2010 literature review on OARs focused specifically on advocacy, the attitudes of scholars and authors, and operational concerns and deployment.[15] Leslie Chan, in her 2004 article, covers the relationship of institutional repositories to the open access movement.[16] Sherrie S. Bergman focused her research on the crisis in academic publishing, citing the increasing cost of scholarly journals,[17] and, like John Willinsky (2003), their impact on academic library budgets and new methods for scholarly communication--including OARs.[18] Alma Swan and Sheridan Brown,[19] as well as Ian Rowland and David Nicholas[20] both track the attitudes of journal authors in relation to open access publishing as compared to their actual behavior.

Justification

Current research uses measurements of quantity, not quality, to assess the value of OARs. In response, our study aims to open up new areas of scholarly inquiry by shifting the focus to the value of an institution's repository. This study will measure the efficacy of institutional repositories in collecting and making available the intellectual output of their highest impact faculty.

Research Questions

Our research attempts to answer the following questions:

1. Do institutional repositories contain the highest impact articles produced by their faculty? If not, do the repositories contain alternate forms of intellectual output from faculty producing the highest impact articles?

2. Is the value of the institutional repository dependent upon institution type (i.e. private not-for-profit versus public institutions as well as high level research universities versus four-year undergraduate colleges)?

3. What kinds of materials are being collected, and does the type of institution impact the kinds of materials being collected?

Method

Sample Selection

We conducted a cross-sectional study of American colleges and universities with open access institutional repositories. The study includes two controls: institution type (private not-for profit or public) and basic classification as defined by the Carnegie Classification System. This classification system, developed by the Carnegie Commission on Higher Education, was selected because it is the most widely accepted description of institutional diversity in the U.S. higher education system.

A sample of forty-eight institutions (n = 48) was chosen from a population of 161 academic institutions listed in the Directory of Open Access Repositories (*Open*DOAR). A leader in repository directories, *Open*DOAR offers a comparably large population, with listings that are quality controlled by trained administrators. In order to select the sample, the population was sorted by institution type and basic classification. Five institutions were selected from each group using an online random number generator. For those groups that did not have five institutions to choose from, all were selected.

Research Design

We wished to measure the value of the institution's repository. To do this, we first consulted Scopus, a citation database, to cull author and title information for the top five most cited articles affiliated with each institution. We then searched the contents of the repositories and recorded the number of highest impact articles available through the institutional repository and the number of highest impact authors that included other items of intellectual output in the repository (i.e. additional papers, syllabi, conference proceedings). *Open*DOAR's quality controlled description of each repository was used to determine the types of materials included.

Limitations

Classifications such as Doctoral Research Universities (DRU) did not contain five repositories for testing, therefore preventing us from collecting a sample size of any significance.

In choosing an appropriate measure for identifying articles and authors of highest impact we decided to use citation indexing despite the criticism this measure has received. For the purposes of this study it was more important to identify scholarly materials that the institution *perceives* as having the highest impact, and in this way citation indexing remains the industry standard.

An additional limitation was the discipline specificity of the chosen citation index. Scopus is the largest abstract and citation database, but it is primarily focused on the sciences. In 2009, it doubled its coverage of the Arts and Humanities when it added 1,450 new titles, but this number is still dwarfed by the 18,000 titles that comprise its holdings. Since the content of OARs is currently dominated by the sciences, we do not expect this will skew the results.

Findings

Of the 48 repositories in the sample, three were discarded because they were either no longer open access or the repository was discovered to be discipline specific and therefore outside our scope. This reduced the sample size to 45 (n = 45). Of these repositories, only three contained the highest impact articles generated by faculty at their institution. Sixteen repositories (36% of the sample) included some form of intellectual output from these same faculty.

As shown in Figure 1 the repositories managed by private not-for-profit colleges and universities held significantly more high-impact articles and high-impact authors than their public counterparts.

As shown in Figure 2, comparing institutions by Carnegie Classification was less illustrative. Higher level research universities (RU/H, RU/VH, DRU) included more high-impact authors, but almost none of their high-impact articles. Paradoxically, the Master's level institutions contained the most high-impact articles in their repositories. Analyzing the data according to these variables, the repositories at institutions with RU/H classification are of the greatest value.

Collapsing the two variables (see Figure 3), it becomes clear that institutions with a RU/H classification and private not-for-profit status contain repositories with greater value than any other classification. If one calculates the average number of articles contained per repository (see Figure 4), institutions with Master's M classification and private not-for-profit status have repositories of the greatest value. However, as discussed in the limitations, the sample of institutions with this particular classification was so small (n = 2) that it negates this result.

Of the 45 institutions, 60% collected articles, 22% collected books, 33% collected conferences proceedings, 2% collected datasets, 2% collected dissertations, 20% collected learning objects, 2%

collected manuscripts, 51% collected multimedia materials, 22% collected special materials (undefined miscellany), 51% collected theses, 18% collected references, and 36% collected unpublished materials. From this information, we know that articles, multimedia, and student theses are collected with the greatest frequency. However, as seen in Figure 5 and Figure 6 there are no significant results to determine what kinds of repositories are collecting what kinds of materials. Public colleges and universities tend to collect a wider array of materials, while private not-for-profit institutions focus on the colleges' intellectual output, in the form of student theses and published articles.
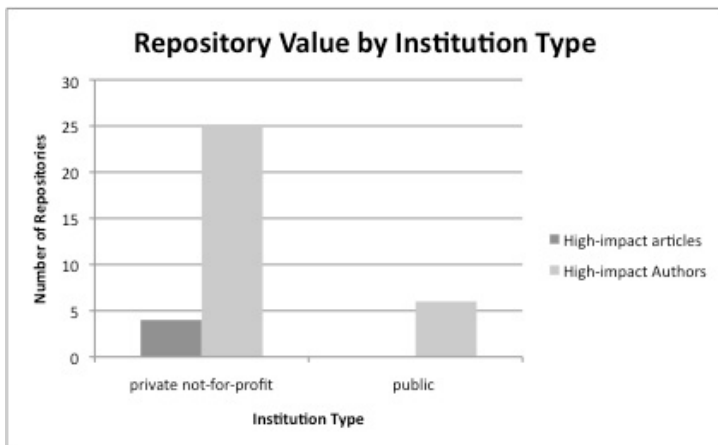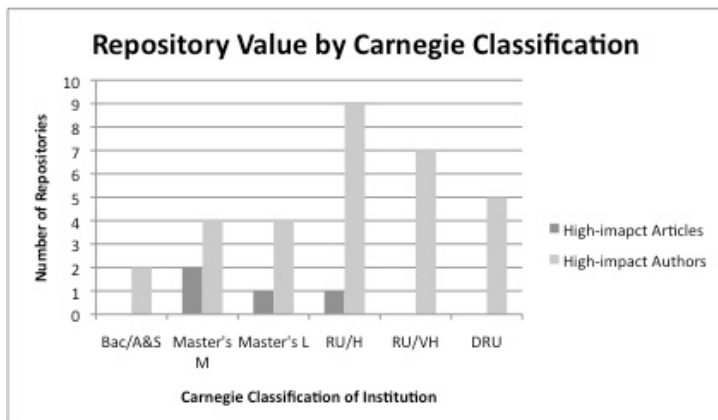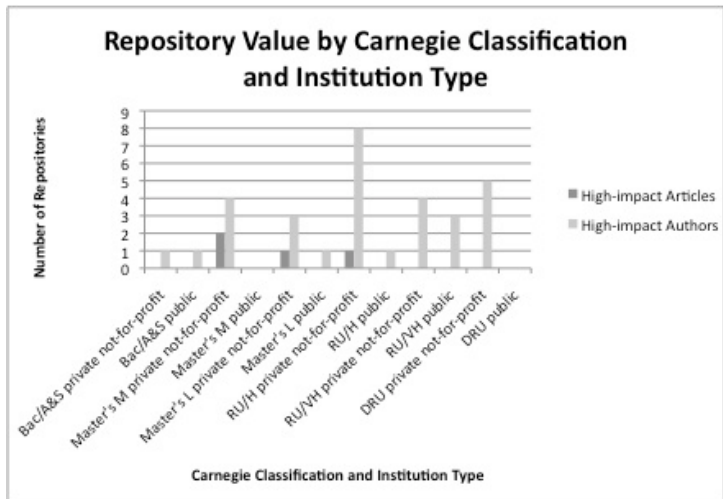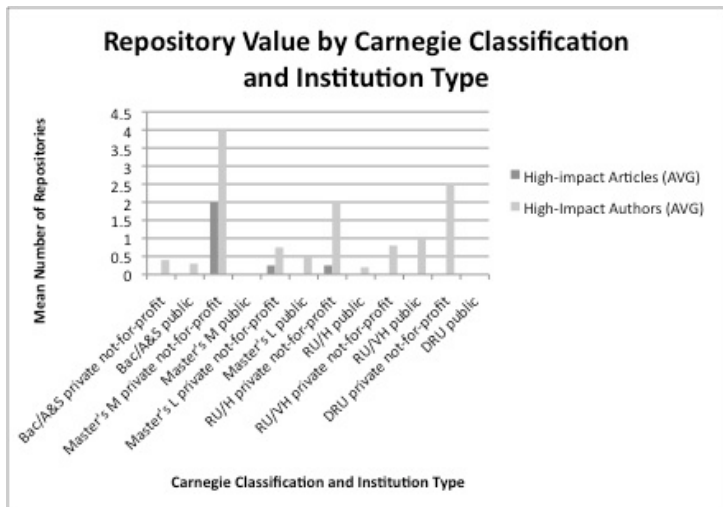

Figure 1

Figure 2
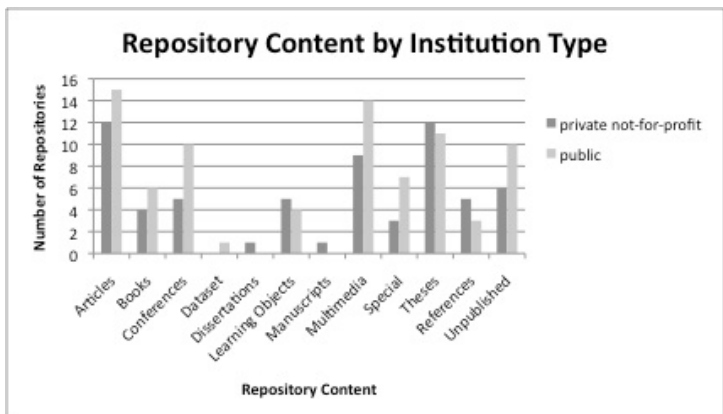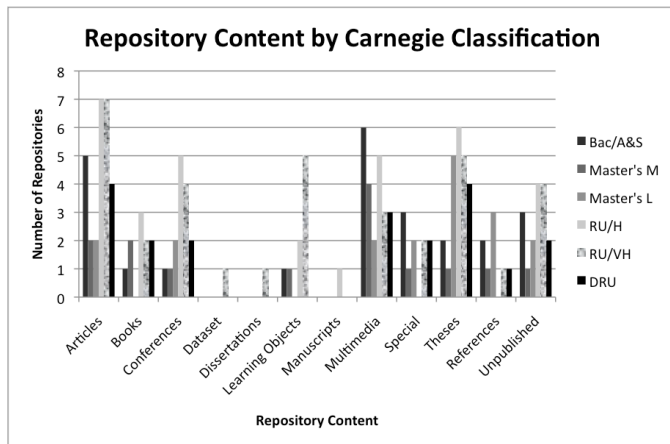
Figure 3

Figure 4

Figure 5

Figure 6

Discussion

While further research is required to attain results of statistical significance, the data suggests that institutions collect according to their missions. For instance, private not-for-profit institutions focused on research collect the products of that research (in the form of peer-reviewed articles) to draw attention to the intellectual output of their faculty. Public institutions also collect peer-reviewed articles as well as unpublished papers, multimedia materials like archival images, and gray literature such as conference proceedings. These materials reflect a commitment not only to the larger scholarly community, but to their regional community as well.

This leads one to ask, if colleges and universities collect materials to support their mission, why are we not finding those items of the greatest impact? As Dorthea Salo states in her 2008 article "Innkeeper at the Roach Motel," the 'build it and they will come' proposition has been decisively proven wrong. Citation advantages and preservation have not attracted faculty participants...."[21] Although our research shows articles as among the most collected materials in the OARs we studied, the quality of those materials—as suggested by the relative absence of high impact articles across institutional type—indicates a lack of commitment on the part of faculty to deposit their best work and the failure of their institutions to collect it. Furthermore, it suggests that in situations where faculty are unable to deposit

9

their work due to conflicts with their publication agreement, their institutions fail to collect related

materials that capture their scholarly output.

Our study points toward the failure of faculty and institutions to make their highest impact articles

available.   However, in conducting this research we began to question the quality of the repository

structure itself, in particular how it was designed to meet users' needs, whether or not the repository

appeared to be maintained, and, most importantly, if the institution was putting real support behind

promoting the OAR's importance to the institution and the community.   Stuart M. Basefsky in his 2009

article argues for an expanded role for institutional repositories. He states: "In building IRs, the evidence

is clear that their mere existence does not translate into use. Hence the necessity to come up with

Harvard-like mandates to force compliance of faculty."[22] Salo furthers this argument by suggesting that

managers of institutional repositories "seek forgiveness rather than permission,"[23] and actively collect

content that is already available on the open Web. The extremes between non-participation and forced

participation broadcast loud and clear that the system is broken, and a new model for institutional

repositories is required.

In their 2009 article, Laurent Romary and Chris Armbruster argue for a more centralized approach to

OAR development, and use the National Institute of Health's 2008 public access policy for publicly

funded research, and the creation of PubMed Central as a compelling example.[24] While subject based

repositories may not be the ideal model for institutional repository development, there is a need for more

meaningful commitment on the part of academe to invest in IRs and redefine their purpose.   Salo,[25]

Marcus Banks (2006),[26] and Hans Roosendaal (2003)[27] all identify gray literature as material well suited

for institutional repositories. As Salo puts it, "Libraries whose support for repositories rests purely on

hopes of collecting peer-reviewed literature would be well-advised not to bother with them."[28]

Basefsky[29] as well as Romary and Armbruster[30] propose that repositories be designed to offer faculty more than just a place to store their output. Features might include easy depositing of materials, a stable system for referencing other faculty's content, exportable data concerning faculty output, and a customizable interface. Romary and Armbruster further suggest "editorial support" as a possible way to add value to repositories.[31] This service could include the pro-active participation of library staff to seek out materials from faculty and track the publishing activities at their institution. Anthony Cocciolo (2009) goes a step further by advocating for a web 2.0 style interface to encourage faculty involvement. PocketKnowledge, instituted at Teachers College at Columbia University, increased user participation in their subject specific repository by 9,728%; with 23% of tenure and non-tenure faculty making at least one contribution.[32] These results, Cocciolo supposes, are the product of a shift in focus away from the needs of the library and toward those of the faculty themselves.[33]

Conclusion

This study found that institutional repositories are not consistently collecting materials of the highest impact, suggesting that the value of these repositories is relatively low. Further research is needed to more fully describe the actual impact of IRs, using indicators that expand upon our findings. As McDowell states, "Although the quality of deposited content is one indicator of successful recruitment, sheer quantity is also necessary for repositories to further the cause of open access. Repository quantity or size is also one of the most tangible, quantifiable mechanisms for evaluation."[34] Future research might seek to combine measures of both the quality and quantity within institutional repositories in order to more accurately reflect their value. Additional studies might also consider alternatives to citation indexes in identifying the materials of the highest impact within repositories, and to provide a possible guide for librarians to target future collection development.

Ultimately, the inconclusive nature of these results indicates that rather than measuring the abundance of marginally functioning repositories, it might be more fruitful to concentrate on alternative models of

OARs that appear to be making headway. Initiatives like PocketKnowledge, as well as the IR developed

at the University of Rochester with the guidance of Librarian, Suzanne Bell and Software Engineer,

Nathan Sarr,[35] show that IR use is not necessarily a thing of the past. In order to fulfill the promise of

open access repositories, the academic community must begin to develop effective methods of evaluation

and innovation in collection development and IR usability.

References

1. Raym Crow, "The Case for Institutional Repositories: A SPARC Position Paper," *The Scholarly Publishing & Academic Resources Coalition* (2002): 1–37.

2. Clifford A. Lynch, "Institutional Repositories: Essential Infrastructure for Scholarship in the Digital Age," *ARL Bimonthly Report* 226 (2003): 1.

3. Clifford A. Lynch and Joan K. Lippincott, "Institutional Repository Deployment in the United States as of Early 2005," *D-Lib Magazine* 11, no. 9 (2005).

4. Cat S. McDowell, "Evaluating Institutional Repository Deployment in American Academe Since Early 2005: Repositories by the Numbers, Part 2," *D-Lib Magazine* 13, no. 9/10 (2007).

5. Charles W. Bailey Jr., Karen Coombs, Jill Emery, Anne Mitchell, Chris Morris, Spencer Simons, and Robert Wright, "Institutional Repositories," *Association of Research Libraries SPEC Kit* 292 (2006): 1–21.

6. Karen Markey, Soo Young Rieh, Beth St. Jean, Jihyun Kim, and Elizabeth Lake, "Census of Institutional Repositories in the United States: MIRACLE Project Research Findings," *Council on Library and Information Resources* (2007): 1–167.

7. Mohammad Hanief Bhat, "Open Access Repositories in Computer Science and Information Technology: An Evaluation," *IFLA Journal* 35, no. 3 (2009): 243–257.

8. Mohammad Hanief Bhat, "Community Engagement in Indian Open Access Repositories: A Deposit Activity Profile," *Chinese Librarianship* no. 27 (2009), http://www.iclc.us/cliej/c127bhat.htm (accessed March 2011).

9. Rowan Brownlee, "Research Data and Repository Metadata: Policy and Technical Issues at the University of Sidney Library," *Cataloging & Classification Quarterly* 47, nos. 3/4 (2009): 370–379.

10. Stevan Harnad, "Southampton Workshop on UK Institutional Open Access Repositories. *Serials Review* 31, no. 2 (2005): 175–177.

11. Blanca Rodriguez Bravo and Maria Luisa Alvite Díez, "E-science and Open Access Repositories in Spain," *OCLC Systems & Services* 23, no. 4 (2007): 363–371.

12. John W. Chapman, David Reynolds, and Sarah A. Shreeves, "Repository Metadata: Approaches and Challenges," *Cataloging & Classification Quarterly* 47, nos. 3/4 (2009): 309–325.

13. Jessica Branco Colati, Robin Dean, and Keith Maull, "Describing Digital Objects: A Tale of Compromise," *Cataloging & Classification Quarterly* 24, nos. 3/4 (2009): 326–369.

14. Jung-Ran Park, "Metadata Quality in Digital Repositories: A Survey of the Current State of the Art," *Cataloging & Classification Quarterly* 24, nos. 3/4 (2009): 213–228.

15. Mohammad Hanief Bhat, "Open Access Repositories: A Review," *Library Philosophy and Practice* 12, no. 2 (2010): 1–8.

16. Leslie Chan, "Supporting and Enhancing Scholarship in the Digital Age: The Role of Open-Access Institutional Repositories," *Canadian Journal of Communication* 29 (2004): 277–300.

17. Sherrie S. Bergman, "The Scholarly Communication Movement: Highlights and Recent Developments," *Collection Building* 25, no. 4 (2006): 108–128.

18. John Willinsky, "The Nine Flavors of Open Access Scholarly Publishing," *Journal of Postgraduate Medicine* 49, no. 3 (2003): 263–267.

19. Alma Swan and Sheridan Brown, "Open Access Self-Archiving: An Author Study," *Key Perspectives Limited* (2005): 1–97.

20. Ian Rowland and David Nicholas, "Scholarly Communication in the Digital Environment: The 2005 Survey of Journal Author Behavior and Attitudes," *Aslib Proceedings* 57, no. 6 (2005): 481–497.

21. Dorothea Salo, "Innkeeper at the Roach Motel," *Library Trends* 57, no. 2 (2008): 98.

22. Stuart M. Basefsky, "The End of Institutional Repositories and the Beginning of Social Academic Service: An Enhanced Role for Libraries," *Conference Proceedings, Presentations, and Speeches*, Paper 7 (2009): 8.

23. Salo, "Roach Motel," 119.

24. Laurent Romary and Chris Armbruster, "Beyond Institutional Repositories," *SSRN* (2009): 1–21.

25. Salo, "Roach Motel," 98–123.

26. Marcus A. Banks, "Toward a Continuum of Scholarship: The Eventual Collapse of the Distinction between Grey and Non-grey Literature," Presented at the Seventh International Conference on Grey Literature, Nancy, France, December 2005.

27. Hans E. Roosendaal, "The Information Market for Research and Higher Education: How to Integrate All Relevant Information in a Network of Repositories," Presented at the Fifth International Conference on Grey Literature, Amsterdam, Netherlands, December 2003.

28. Salo, "Roach Motel," 119.

29. Basefsky, "The End of Institutional Repositories," 1–13.

30. Romary and Armbruster, "Beyond Institutional Repositories," 1–21.

31. Ibid., 12–14.

32. Anthony Cocciolo, "Can Web 2.0 Enhance Community Participation in an Institutional Repository? The Case for PocketKnowledge at Teachers College, Columbia University," The *Journal of Academic Librarianship 36*, no. 4 (2010): 310.

33. Cocciolo, "Can Web 2.0 Enhance Community Participation," 312.

34. McDowell, "Evaluating Institutional Repository Deployment," under "Size of Repositories."

35. Suzanne Bell and Nathan Sarr, "Case Study: Re-engineering an Institutional Repository to Engage Users," *New Review of Academic Librarianship* 16, no. 1 (2010): 77–89.