AC 2010-736: OPEN ACCESS AVAILABILITY OF PUBLICATIONS OF FACULTY IN THREE ENGINEERING DISCIPLINES

Virginia Baldwin, University of Nebraska, Lincoln

Professor Virginia (Ginny) Baldwin is the Engineering, Physics, and Patent and Trademark Librarian at University of Nebraska - Lincoln. She received her MLS Degree from Indiana University in 1990. Her BS in Mathematics is from the University of North Carolina, and her MS in Mathematics is from Florida Institute of Technology. She is a member of the Engineering Libraries Division and is currently serving as Past Chair of the Science & Technology Section of the Association of College & Research Libraries Division of the American Libraries Association.
Open Access Availability of Publications of Faculty in Three Engineering Disciplines

Abstract

The analysis presented here provides a snapshot in time of the open access online availability of the five most recent works of engineering faculty in five institutions that overall have heavily populated institutional repositories. The incentive for the study was to provide a measure of the inclination of engineering faculty in specific disciplines to provide open access to their most current manuscripts or articles. The Web of Science database was used to choose the five most recent publications for each faculty member in each of three disciplines: civil, chemical, and mechanical engineering. The study was done for the five United States universities that had the most overall content in their repositories, as ranked by the Registry of Open Access Repositories (http://roar.eprints.org) at the time of data collection. One of these universities was ranked by U.S. News & World Report as having the number one engineering college among schools whose highest degree is a doctorate. To determine open access availability of each article via deposition in an institutional repository, each publication title was searched using the ROAR Content Search interface. To determine other forms of open access a Google Scholar search was used. The results of an analysis of these data collected for various sources of open access with breakdowns by college and department are provided.

Introduction

Under the leadership of MIT in developing the DSpace system, institutional repositories (IRs) [long term electronic archives of works authored by affiliates of the institution] emerged in the United States in the fall of 2002\(^1\). Built on the concept of open access, defined for these purposes as freely available online access to full text articles, and with discipline repositories such as arXiv, as a precursor, many were founded with the “build it and they will come” concept.

The Scholarly Publishing and Academic Resources Coalition (SPARC)\(^2\), whose activities are meant in part to serve as an aid in reducing financial pressures on libraries, defines institutional repositories in terms of capturing and preserving intellectual output in their August 2002 position paper\(^3\). The paper goes on to stress the importance of open or low-barrier access and the ability to share metadata with external systems to facilitate access to the broader research community.

Clifford Lynch, Executive Director, Coalition for Networked Information, defines IRs in terms of a set of management and dissemination services provided by institutions. He stresses the importance of managing technological changes by migrating digital content to assure that at least their own intellectual content is preserved. With libraries increasingly cancelling their print subscriptions in favor of somewhat less costly online only access, where are the assurances to faculty that back content will be available to them through technology changes, publisher mergers, future journal cancellations, and whatever else might transpire to cause loss of access to online content? Lynch goes on to state “that a primary responsibility of our universities is both to make these resources available and to preserve them.”\(^4\) Touting the advantages of their own IR software platform, the Digital Commons, the Berkeley Electronic Press (BePress) President and Director of Journals explains that their product expands upon the service aspect, to actually
impacting open access publishing. The concept behind their platform is institutional customization that transforms the IR from an archival vault to a way for faculty to organize and showcase their scholarly work.

In addition to preservation and stewardship there is another concept that IRs address, that of “fundamentally changing the way scholars disseminate their research.” Lynch addresses this aspect in an interview and states that open access is not about saving libraries money, desirable and welcome as this might be. Rather “the first and most central point of the Open Access movement is to facilitate the growth and dissemination and use of knowledge and scholarship by removing barriers and friction.” Law Library Director and Associate Professor of Law Carol A. Parker relates IRs to open access. She states that more efficient sharing of their work through self-archiving in IRs enables law school faculties to receive valuable prepublication feedback and to reach new audiences with the potential to redefine scholarly publishing in the legal field. “At a minimum,” she writes, self-archiving scholarship increases “the impact of the work through expanded readership and faster access.”

Populating IRs is a frequent topic in the relevant literature. Marketing and branding are stressed and predominate. Where and how library web sites link to the IR, how (and whether) they address the issue of scholarly communication, and terminology are all important. Terms for IR access points such as IU-Bloomington’s IUScholarWorks and University of Oregon’s Scholars’ Bank on the library homepage connote the faculty scholarship rather than the institutional aspect of IRs and are therefore deemed more inviting to faculty. Indeed the current number one ranked IR in terms of total number of records on the Registry of Open Access (ROAR) site is Brigham Young University (BYU). BYU’s Scholar’s Archive, their IR name, is an example of terminology that may be more inviting to faculty. In a comparison of deposition by subject area, Xia counted the number of deposits by discipline for seven universities: three British, one Swedish and three Australian to obtain rates of deposition in the fields of chemistry, physics, economics, and sociology and presented the results as percentages (depositing rates). Two of the universities had considerably higher rates of deposition, (one in chemistry, one in physics), than the other disciplines and the other universities when weighted to account for the varying number of faculty at the institutions. Jantz and Wilson studied the IRs of ARL academic institutions in eight disciplines and reported relatively high number of deposits of faculty works in the field of Engineering (650), with Bioscience (1081), Math (1414), and Economics (1090) higher. It would have been interesting to know how much of the material deposited in these areas was very recent. Based on the Census of Institutional Repositories in the United States, Smith reported that less representation would be expected from masters or baccalaureate institutions.

As IR deposition grows it will be increasingly important to measure the research impact of deposited items. Library directors and other library representatives in academic institutions at various levels of IR deployment rated 16 anticipated benefits of IRs in a survey reported in an unpublished paper. One of the top-ranked benefits was “Exposing your institution’s intellectual output to researchers around the world who would not otherwise have access to it through traditional channels.” Publications of research in foreign countries are often cited by engineering faculty. The Registry of Open Access Repositories (ROAR) indexes repositories all over the world and provides statistics on foreign countries’ total records as well.
There is ample justification for the works of scientists and engineers to be freely available online, no more true than in the field of medicine. No less than the United States Congress has gotten involved by passing the NIH Public Access Policy that mandates public access to final peer reviewed manuscripts from NIH funded research within 12 months of publication. The Swiss National Science Foundation (SNSF)\textsuperscript{14} signed the Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities\textsuperscript{15}. The \textit{Journal of Biological Chemistry (JBC)} published on-line in 1995 by the American Society for Biochemistry and Molecular Biology (ASBMB) and its partner, Stanford University's HighWire Press released the back issues of \textit{JBC} on-line free to anyone with Internet access and initiated \textit{JBC} Papers in Press which releases all accepted \textit{JBC} papers on the day they are accepted\textsuperscript{16}. In the article “Effective Web Dissemination of Construction IT Research Publications,” Civil and Geodetic Engineering Professor Turk and Economics and Business Administration Professor Bjork state, “Open access provides increased readership within the research community but perhaps more importantly it increases readership by industry experts” and in fields such as civil engineering can thus have an impact on practice\textsuperscript{17}.

Jantz and Wilson analyzed IR deposits in selected subject areas in 49 ARL Libraries and found that 660 out of 5,140 items deposited (13%) were from authors in the field of engineering. Interestingly they state that 60% of all represented work comes from one institution\textsuperscript{8}. A similar result occurred for the Zuber study\textsuperscript{18}. Of the 41 (not necessarily ARL) institutions studied, one institution, University of Michigan, accounts for almost half of the total holdings in the study. A scan of the names of the institutions used in the study reveals that University of Michigan is the only one that is also in the “top five” used in the current study. The overall results for the Zuber study showed a distinct advantage in the engineering disciplines, most heavily in mechanical and electrical engineering, over the others represented. Xia analyzed the IR deposits of physicists at an institution with an IR deposit mandate policy, the University of Southamption in England, and compared faculty deposition in their IR with their deposition in the subject repository arXiv. Xia concluded from this small sample that there appears to be hesitancy for physics authors to make their articles repeatedly available in a second repository\textsuperscript{11}.

\textbf{Methodology}

The analysis presented here provides a snapshot in time of the open access online availability of the five most recent works of engineering faculty in five institutions that overall have heavily populated institutional repositories. The incentive for the study was to provide a measure of the inclination of engineering faculty in specific disciplines to provide open access to their most current manuscripts or articles.

Among the caveats that deserve recognition are the following:

1. Overall content amount in the IR does not imply high content for the Engineering Faculty of the institution.
2. If an author submits the article to an open access journal or to PubMedCentral (PMC), how important then is deposition by the author in his/her institution’s IR?

This study used only the works of tenured/tenure track faculty whose names were obtained in the spring of 2008 from the University Web sites of their affiliated engineering colleges or schools.
Excluded from the study were research, retired, and emeritus faculty and lecturers. The study used three departments in common to all, Chemical or Chemical and Biomolecular Engineering (chemical engineering), Mechanical Engineering, and Civil or Civil and Environmental Engineering (civil engineering). For the institution for which the word “environmental” was not in the department name (University of Nebraska-Lincoln), the study of environmental engineering was one of the areas of concentration for the department.

The author used the Thompson–Reuters database, *Web of Science (WOS)*, limited the search for each faculty member to their current institution, and selected their five most recent citations as of summer of 2008. For some faculty members one or more of the citations recorded were from their presence in some other department at their institution. Since these were included in the study, there is some representation from disciplines other than the three departments. Of course this is very much true anyway as considerable interdisciplinary work is represented in these citations. Authors of works were cross-checked within departments to remove duplicates, but no cross checking was done from department to department. Because WOS is the only science/engineering multidisciplinary article index whose institutional subscription was available to the author, the publications on which this study is based are those journals that it covered. Deposition of another important publication type, conference proceedings, could not be included.

The Registry of Open Access Repositories (ROAR) was established to provide information about several types of open access repositories throughout the world as well as federated searching of their content. The ROAR Web site provides statistics related to repository content. Among these statistics is a ranking of repositories by content amount “Total Records.” For IRs in the United States the top five IRs, as of March 6, 2008, are the ones used in this study. Their number of records, software used, and dates of registration are given in Table 1. (As of early 2010 Brigham Young University and Texas A&M University have moved into the number one and two positions, with MIT and Georgia Tech now dropped out of the top five). MIT, used in this study, has the *U.S. News & World Report* number one ranked Engineering School for graduates in the United States while those of two others, Georgia Tech and University of Michigan, are among the top ten. For undergraduate education in mechanical and civil engineering, these three are in the top ten and two of them are in the top ten in chemical engineering. These rankings and their status as top five IRs provided a rationale for their use in this study.
Table 1 IR Data for the Top 5 IRs in Number of Records as of March 2008

To determine online availability of the faculty publications, two search interfaces were used. The ROAR site has a “Content Search” feature and recognizes phrases when quotation marks are used around the words in the phrase. “Content Search” uses a Google Scholar Custom Search Engine to search for the article in all of its registered sites. A second search was done using Google Scholar, again with the title or part of the title in quotes, recognized by Google Scholar as a phrase, since it retrieves other forms of open access availability as well. The study was done from outside the IP range of the author’s institution so there was no confusion about whether access to full text resulted from affiliated subscriptions. The interest of this research was to determine whether the engineering faculty author (or his/her institution through harvesting) had provided open access to his/her publication, therefore other forms of online availability provided by a non-author were not recorded. The following were tracked, in this order:

1. The article was available in the IR of the author’s institution.
2. The article was available in the IR of a co-author’s institution or was deposited in a disciplinary open access repository.
3. The article was available by the author posting it on his/her Website or the Website of a research group or department at the author’s institution.
4. The article was available by a co-author posting it on his/her Website or the Website of a research group or department at the co-author’s institution.
5. The article was available on PubMedCentral (PMC) and/or on some other government Web site (also other than that of a co-author).
6. The article was available directly from the publisher either because it was in an open access journal or some factor caused the article to be open access at the journal’s Web site.

In most cases only one URL was recorded, corresponding to number 1, 2, 3, 4, 5, or 6 above, the lower number representing the highest priority to record. When the article was also available in PMC or a government Web site a second URL was recorded. The article “Using Google Scholar to Search for Online Availability of a Cited Article in Engineering Disciplines” describes some of the difficulties encountered when using an exact article title obtained from Web of Science as a search phrase in Google Scholar and how to recognize and accommodate those difficulties. If
the title was short or only contained common words the year and/or author were added to narrow the search.

**Analysis and Results**

The total number of faculty articles found in each type of open access was counted for each discipline and each institution. Figures 1 through 3 show the corresponding percentages of articles for chemical and biomolecular, civil and environmental, and mechanical engineering disciplines. Note that the range of values on the vertical axis in the chemical and biomolecular engineering chart is nearly three times greater than on the other two charts. First and foremost, the percentage in types 1 and 2 are disappointing, with type 2 (we could call it collaborative archiving) being almost nonexistent. The highest percentages of availability in general are for type 3, a result of the author making it openly available on a faculty or group Web site. Larger percentages for type 5, PMC or government Web site access, occur for chemical and civil engineering than for mechanical, possibly because of a higher level of NIH grant funding in those areas that makes deposition mandatory within 12 months of publication. This study was done on articles published in the summer of 2008 and the searches for open access availability were all conducted more than 12 months after the citation data for the articles was collected. Another percentage that stands out in the charts is the high percentage of IR deposition in the field of chemical engineering at the University of Nebraska-Lincoln. The reason for that is explained in the information obtained by interviewing Paul Royster, highlighted in the subsection below.

![Figure 1 Percentage of Open Access by Type for Chemical and Biomolecular Engineering](image)

**Figure 1 Percentage of Open Access by Type for Chemical and Biomolecular Engineering**
Figure 2 Percentage of Open Access by Type for Civil and Environmental Engineering

Figure 3 Percentage of Open Access by Type for Mechanical Engineering

<table>
<thead>
<tr>
<th>Department/Open Access Type</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chemical Engineering</td>
<td>220</td>
<td>50</td>
<td>9</td>
<td>51</td>
<td>10</td>
<td>9</td>
<td>11</td>
</tr>
<tr>
<td>Civil Engineering</td>
<td>310</td>
<td>22</td>
<td>1</td>
<td>70</td>
<td>17</td>
<td>16</td>
<td>20</td>
</tr>
<tr>
<td>Mechanical Engineering</td>
<td>382</td>
<td>17</td>
<td>14</td>
<td>53</td>
<td>25</td>
<td>7</td>
<td>9</td>
</tr>
<tr>
<td>Total</td>
<td>912</td>
<td>89</td>
<td>24</td>
<td>174</td>
<td>52</td>
<td>32</td>
<td>40</td>
</tr>
<tr>
<td>Percentage of Total</td>
<td>69%</td>
<td>7%</td>
<td>2%</td>
<td>13%</td>
<td>4%</td>
<td>2%</td>
<td>3%</td>
</tr>
</tbody>
</table>

Table 2 Count of Open Access by Type and Discipline, All Institutions Combined
Table 2 gives a tabulation of each of the 6 types of open access (with 0 representing those articles with no open access) totaled for all institutions in each discipline. For clarity, the tabulation is such that, for example, the 40 articles that were found open access at the publisher’s Web site (type 6) were not found open access in types 1 through 5, the 32 that were found open access in PMC or a government Web site (type 5) were not found open access in types 1 through 4, etc., priority being given to the lowest type number when more than one type of open access occurred for an article. In Table 2 the percentages for each type of deposition are given for all data combined across departments. Types 1, 2, and 5, a total of 11%, all represent a form of archival deposition for the most part. The total percentage of these types plus type 6, open access at the publisher’s Website, is 14%, a small, though not insignificant percentage. Figure 4 is a chart of the count for each of types 1-6, all institutions combined, from Table 2. Another way to use the data in this study to assess the inclination of faculty in these departments to provide open access is by compiling the number of faculty who through their own or their co-author’s actions, had at least one article that was freely available online. These numbers tell a remarkably different story as seen in Figure 5, with all but three at 50% or greater. What makes these numbers so much higher is the considerable number of faculty for whom few (and often only one) of their articles are freely available online in this manner. There are numerous possible explanations that involve timing, publisher restrictions, greater deemed or actual significance of a particular article, etc. In Figure 5 as in the Figures 1-3, the University of Nebraska Chemical and Biomolecular Engineering Faculty have the highest percentage. Conversations with three librarians, each in charge of their institution’s repository and ultimately for obtaining content, is presented in the next section and explains some of the differences in the charts and tables.

![Figure 4 Count of Open Access by Type and Discipline All Institutions Combined](image-url)
Conversations with Librarians from Three Surveyed Institutions

The following is a summary of relevant information obtained by phone or email conversations with Librarians from three of the five institutions. Each of these librarians has a position of responsibility for their institution’s repository.

Paul Royster, University of Nebraska-Lincoln (UNL), Coordinator for Scholarly Communication, provided many insights. In this dedicated position from UNL Digital Commons’ infancy, Paul has part time support from a second librarian who identifies works to be posted and undergraduate student workers who scan and upload UNL author or co-author works. UNL theses and dissertations that are open access in the Digital Commons are placed there separately from their appearance in the main repository section that is allocated to this type of content and is not open access. For engineering and science Paul reports that by far the most cumbersome issue is manuscripts that include tables, graphs, equations, sometimes even Greek characters that are inset or provided separately in graphic format and make scanning and checking the scans of the document often prohibitively time consuming. This is only a necessity, of course, for those publishers who do not permit the posting of the final published PDF format. Paul has an easy solution to that – encourage faculty to be selective in where they publish. Another solution is to coordinate research with a government employee for whom no copyright restrictions are allowed. He cites three publishers, all organizations, which allow the posting of the published work: Institute of Electrical and Electronics Engineers, American Physical Society, and the American Institute of Physics. There are many others, of course, and Paul uses the RoMEO service on the SHERPA Website as a guide. This service monitors publisher copyright and archiving policies and differentiates the policies by color. The following is the ROMEO colour Archiving policy (http://www.sherpa.ac.uk/romeoinfo.html#colours):

- **green** can archive pre-print and post-print or publisher's version/PDF
- **blue** can archive post-print (i.e. final draft post-refereeing) or publisher's version/PDF
yellow can archive pre-print (i.e. pre-refereeing)
white archiving not formally supported.

For the fields of science and engineering Paul concentrates on the SHERPA “green” publishers when using an author’s CV or Web site to identify publications. The Department of Chemical & Biomolecular Engineering at UNL hired graduate students to assure populating the Digital Commons with a large portion of its faculty publications. This explains the exceptionally high percentage of deposition for their department (>60%) in this study. Additional information about the UNL Digital Commons is available in it at http://digitalcommons.unl.edu/cgi/viewcontent.cgi?article=1063&context=libraryscience.23

Jim Ottaviani, University of Michigan Library, Coordinator, Deep Blue, reports that their repository, Deep Blue, opened and was formally announced with a substantial body of work in it already, over 20,000 items. The library had and still has a robust technical and content infrastructure in place to enable the harvesting of batch loads of available materials. In renewing licenses with several restrictive publishers, the university has been able to negotiate author rights back to the university to allow the deposition of published material, though often this only applies to earlier work. Thus they have been able to harvest earlier author publications for deposition without getting permission from the author, who, often surprisingly to him/her, gave those rights away years or days ago. University of Michigan Libraries, as well as UNL Libraries offer copyright support service to help faculty determine their deposition rights. UM also offers a format blind policy, so that along with the paper the author is able to include data, videos, notes, etc, all within a single permanent, citable URL.

Sara Fuchs, Georgia Institute of Technology (GT) Digital Initiatives Librarian, Scholarly Communication & Digital Services, came into this position two years after their repository, SMARTech, was launched. Having to deal almost entirely with the graphics issue, Georgia Tech being an Engineering school, GT initially populated their repository by harvesting material already on departmental Web sites. Most of what was obtained through this process was “grey” literature such as technical papers, working papers, white papers, and some journal articles, (the latter posted only after checking with SHERPA policies). The grey literature is an important body of work to archive. When a publisher’s policy is unknown GT often contacts the publisher and usually is able to obtain permission, especially with the smaller publishers. Half of the material in the GT repository is theses and dissertations. When graduate students ask for their own theses or dissertations that they cannot access online, GT is then generally able to obtain the student’s permission to make it available. One of the services offered is conference uploading software, and this material is added to the repository as well.

Conclusions

What nearly assures an expansive and continual increase in the amount of science and engineering literature that is freely available online regardless of institutional subscription ownership is compounded by the:
1. increased interdisciplinary nature of research, government mandated access to research produced by NIH grants, and the possibility of expansion to other funding agencies of the government,
2. proliferation of research groups comprised of faculty and industry researchers throughout the world and
3. ease with which faculty can post their publications to the Internet and make them freely available.

The continuing advancement of search interfaces and the development of search interfaces dedicated as finding aids to institutional and disciplinary repositories assures expanded intellectual access as well.

The results of this project support these conclusions. In this study nearly 11% of the articles were deposited in institutional or disciplinary repositories. Much higher were the percentages of faculty in all three disciplines, in all five institutions who have at least one article with some form of open access. IR deposition is still in its infancy in these fields and repeated studies such as this one will determine if the percentage of available publications is on the rise. Free access is increasingly becoming the mantra and predictably faculty will more and more prefer to pull their cited publications from the Internet rather than their file cabinets.

Several specific points can be made as a result of this study:

1. Now is the time for engineering librarians to promote the use of the interfaces, Google Scholar and ROAR, to locate articles online that are not available through their institution’s subscriptions. Increased active use of these tools to download needed articles could foster a desire to reciprocate.
2. To promote adherence to copyright law, librarians can encourage faculty to restrict access to online communities that are used to share articles within their research groups. Use an appropriate institutional or disciplinary repository to make available articles relevant to the group’s study and authored by a research group member.
3. The prevalence of branded (with publisher information) PDF versions facilitates determination of whether an online article is the final published one. Otherwise the online version of an article can be evaluated to determine its potential usefulness and a decision can be made on whether or not to order the published article through InterLibrary loan or to buy it from the publisher.
4. Promote a second step to the current version of the “Invisible College.” Encourage faculty to deposit their work in their institution’s IR whenever a copy is requested by a colleague either in addition to or instead of emailing it to the colleague.
5. In the fields of science and engineering, not being able to post the final published document in an author’s IR due to publisher restrictions is a significant barrier. It can be important to advise our faculty of these restrictions and their implications in our conversations with them. “Knowing strong and weak contributors by discipline can assist institutional repository sponsors to better formulate recruitment strategies and generate incentives as a means to increase contribution levels. In addition, identifying institutional repositories that demonstrate strong contribution levels in various academic disciplines can provide an important resource to an IR sponsor who needs advice in improving their
Precisely what has been identified in this study is the role a dedicated repository or scholarly communications librarian, hired student assistant, departmental support, negotiating licenses, and other factors can have in accommodating content deposition. The > 60% rate of deposition for the UNL Chemical and Biomolecular Engineering Department is an encouraging sign for the possibility of greater IR deposition in the future.

6. Finally, studies such as these can be replicated in other fields and disciplines.

Bibliography


