Faculty Self-Archiving: The Gap between Opportunity and Practice

Denise Troll Covey
Principal Librarian for Special Projects
Carnegie Mellon University Libraries

Carnegie Mellon's vision of the universal digital library is free-to-read access to the cultural and intellectual heritage of humankind, including traditional and innovative scholarly work. The success of initiatives to incorporate free-to-read versions of scholarly publications in the digital library hinges on the participation of faculty authors. Authors must retain the right to self-archive their work or publish in open access journals. Despite the availability of publisher self-archiving policies, substantial investments in software and protocol development, and research confirming the impact advantage of open access, faculty reluctance to publish in open access journals is very high and faculty self-archiving practice remains quite low.

To improve our understanding of the scholarly communication practices of campus faculty, in 2006 Carnegie Mellon University Libraries interviewed a stratified random sample of the faculty. The study revealed that only about half of the faculty were aware of efforts to make scholarly work freely available on the web and suggested that only 22% were self-archiving any of their work. The study confirmed our perception that faculty behavior is driven in large part by their focus on promotion and tenure and related concerns about peer review and peer practice. We learned that only about a third of the faculty considers copyright transfer terms when selecting a publisher, and that many do not understand or keep copies of their copyright transfer agreements. Of the few who had tried to negotiate a copyright transfer agreement, their primary concern was the right to re-use their work; only 3% were interested in the right to self-archive their work.[1] Based on the findings from this study, the University Libraries and the Office of Legal Counsel began an ongoing program on Authors' Rights and Wrongs to help faculty understand the economic and access issues and the importance of managing their copyrights.

To improve our understanding of the self-archiving practice of campus faculty and the opportunity to self-archive in different disciplines, Carnegie Mellon University Libraries is currently conducting a study of faculty publication lists available on the web. In phase I of the study, publication lists are being analyzed to identify publication type and access type. In phase II of the study, all of the journal publications are being analyzed to further determine whether the work was or could have been self-archived in compliance with publisher policy. This paper explores preliminary findings from this ongoing study.

Method

Faculty publication lists can be discovered by a variety of means, for example, by searching the university web site or searching Google Scholar. These methods, however, are problematic because of common names, nicknames, and other inefficiencies in terms of recall and precision. After some discussion the Libraries Council agreed that the study would focus on publication lists found by starting with the department home page, following the link to the faculty directory and from there exploring every link associated with each faculty member. We understand that

this will not capture all faculty publication lists or self-archived works, but it will provide a clearer picture of faculty self-archiving practice than we currently have.

Many faculty members are affiliated with multiple departments or centers in the university. To ensure that the study reveals as accurate a picture as possible of self-archiving practice in the university, each faculty member and his or her publications must be counted only once. We agreed to assign faculty and their publications to the faculty member's home department.

With feedback from the Libraries Council and ongoing guidance from the Dean of Libraries, we agreed to the following publication and access codes:

Publication types

- Journal articles
- Conference papers – including conferences, symposia, workshops, annual association or society meetings, etc.
- Technical reports – including working papers
- Book chapters
- Books – including theses and dissertations
- Other – including encyclopedia articles, book reviews, editorials, opinion pieces, testimonials, and works that we could not identify as fitting any other publication type

Access types

- Open access – link to a freely available copy of the full text work
- Restricted access – link to a copy of the full text work available by subscription
- E-commerce – link to a commercial web site with a shopping cart
- Broken link – including file not found or corrupted file
- No link
- On request – email link to author

Phase I of the study entailed identifying the faculty for whom a given department is their home department, locating and printing all of the faculty publication lists and then coding the publication and access types. We quickly discovered that most faculty members have multiple publication lists, many of which contain redundant citations. We agreed that each unique citation should be coded and counted only once, so redundancies had to be eliminated. The task was complicated because of our goal of understanding self-archiving practice and because of the apparent disorganization of faculty publication lists. Some lists had clearly been abandoned. Other lists were not only redundant, but inconsistent about which list provided links to open access full-text copies of the publications. In eliminating redundancies, care had to be taken not to eliminate the occurrence of the citation with the open access link.

Three librarians – Kristin Heath, Diane Covington, and I – are doing Phase I coding and data entry. As leader of the project, I create a spreadsheet for each department that provides a structure for entering the data for each faculty member's citations per publication and access type. When Heath and Covington finish coding Phase I for a department, they send me the completed spreadsheet and give me the printed and coded publication lists. I then create a separate spreadsheet and do the Phase II coding.

Phase II of the study focuses entirely on journal articles. The Phase II spreadsheet lists the titles of all the journals in which the faculty in a given department published and the access type for each publication. After these data are compiled, I identify the publisher and, if possible, the publisher's policy regarding open access for each journal title. This enables an analysis of both the opportunity to self-archive in each department, based on the journals in which the faculty has published and the publisher's open access policy, and the practice of self-archiving in compliance with or breach of publisher policy.

Assessment of compliance in this study is based on two factors:

- Whether the publisher allows or prohibits self-archiving
- Whether the publisher allows, prohibits or requires the publisher version of the work to be self-archived

Compliance with publisher policies regarding embargo periods could not be determined because more often than not the date of self-archiving is unknown. In many, perhaps most cases, distinguishing between pre-print and post-print author versions is not possible. The work required to assess compliance with publisher policy details like the specific text to be displayed before and after publication, removal of pre-prints after publication, links to publisher web sites, etc. was determined to be too difficult or labor intensive to pursue in the study. Doing this work would inordinately lengthen the time it takes to complete the project. The time investment is unwarranted because publishers appear not to be monitoring compliance with their policies. However, basing assessment of compliance on only two policy factors means that the determination of non-compliance in this study is conservatively low. In reality, non-compliance is likely much higher than the findings presented in this study.

Because departments vary significantly in size (the number of faculty) and disciplines vary significantly in the type and volume of publications produced, all data in the study have been converted to percentages for the purpose of comparison.

Preliminary Findings

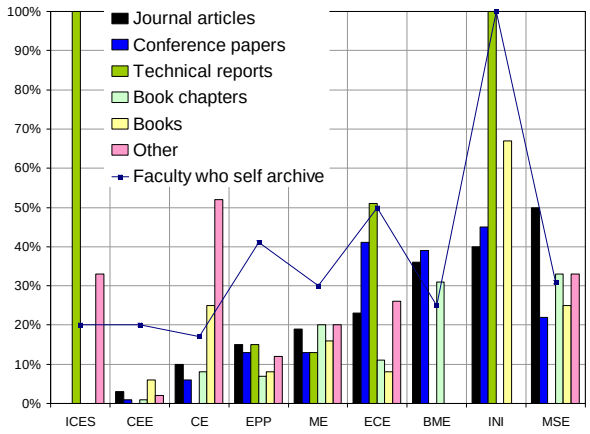
To date, Phase I and II of the study have been completed for the nine departments and research institutes in Carnegie Institute of Technology, which is the school of engineering. Phase I and II have also been completed for three of the seven departments and institutes in the School of Computer Science.

<u>Carnegie Institute of Technology (CIT)</u>

The engineering faculty appears to publish almost as many conference papers as journal articles. They publish ten times as many journal articles as technical reports, fewer book chapters than technical reports, and few books or assorted other materials. Overall, technical reports and conference papers are the most likely works to be self-archived. A third of the technical reports and 31% of the conference papers cited on CIT faculty publication lists have been self-archived. Few books or book chapters (11% each) are self-archived. Other publications such as editorials,

book reviews, and encyclopedia articles were self-archived almost twice as frequently as books or book chapters. Only 24% of the journal articles published by CIT faculty have been self-archived. In contrast publisher policy would allow self-archiving of 67% of the journal articles published by CIT faculty and cited on publication lists accessible from the department home page.

Faculty behavior of course varies across disciplines. Figure 1 shows the self-archiving practice of faculty in the departments and research institutes in CIT. Left to right in Figure 1, the data per department are organized by increasing self-archiving of journal articles. At 50%, Materials Science and Engineering (MSE) faculty self-archived the largest percentage of journal articles, followed by faculty in the Information Networking Institute (INI) and faculty in Biomedical Engineering (BME). INI, BME and Electrical and Computer Engineering (ECE) faculty self-archived the largest percentage of conference papers. Faculty in the two research institutes, INI and the Institute for Complex Engineered Systems (ICES), self-archived the largest percentage of technical reports, followed by faculty in ECE. Faculty in Civil and Environmental Engineering (CEE) and Chemical Engineering (CE) self-archived the least material.



	DEPARTMENT OR INSTITUTE	FACULTY
ICES	Institute for Complex Engineered Systems	10
CEE	Civil and Environmental Engineering	20
CE	Chemical Engineering	23
EPP	Engineering and Public Policy	17
ME	Mechanical Engineering	23
ECE	Electrical and Computer Engineering	60
BME	Biomedical Engineering	8
INI	Information Networking Institute	3
MSE	Materials Science and Engineering	16

Figure 1: Self-archiving practice in CIT, the school of engineering.

Just as faculty behavior varies across disciplines, so does publisher policy. Figure 2 provides data about publisher open access policies of the journals in which engineering faculty published. Of the policies located, most publishers allow self-archiving, though the range is significant, from a low of 54% in Civil and Environmental Engineering (CEE) to a high of 90% for the Information Networking Institute (INI). Chemical Engineering (CE) and Mechanical Engineering (ME) have the highest percentage of publisher policies that prohibit self-archiving

(30% and 29% respectively). No policy located prohibited self-archiving in Electrical and Computer Engineering (ECE).

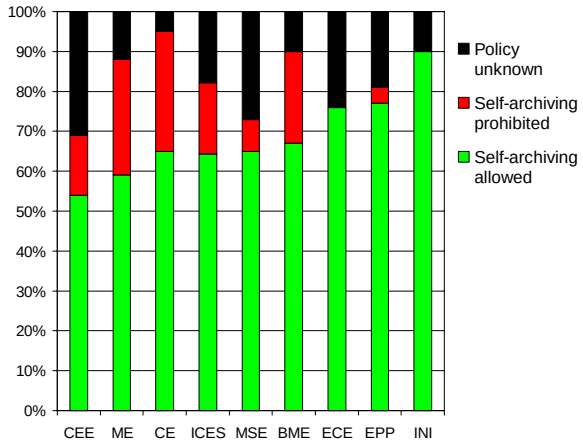


Figure 2: Journal publisher self-archiving policies in engineering.

Figure 3 provides details about the opportunity to and practice of self-archiving in the school of engineering. Left to right in Figure 3, the data per department are organized by increasing opportunity to self-archive journal articles based on the policies of the publishers of the journals in which the faculty published. The Figure also shows the percentage of faculty who self-archive journal articles in comparison with the percentage of faculty who self-archive any type of publication. In most cases, fewer faculty self-archive journal articles than self-archive other types of publications.

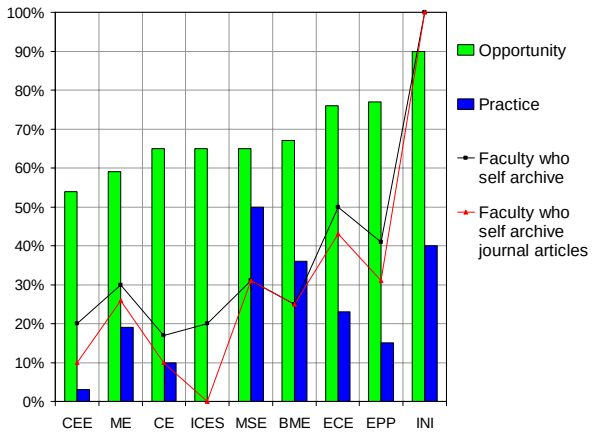


Figure 3: Opportunity to self-archive and practice of self-archiving journal articles in CIT.

There appears to be no correlation between the opportunity to self-archive journal articles in a given engineering discipline and faculty practice. For example, 77% of the journals in which faculty in the department of Engineering and Public Policy (EPP) publish allow self-archiving, but the faculty has self-archived only 15% of their journal articles. Over half of the journals in which Civil and Environmental Engineering (CEE) faculty publish allow self-archiving, but the faculty has self-archived only 3% of their articles.

The data suggest that at least some engineering faculty are not aware of or do not care about publisher policies. See Figure 4. In all departments and research institutes where journal articles have been self-archived the faculty has breached some publisher policies. In three cases most of what is self-archived does not comply with publisher policy. In Mechanical Engineering (ME) 71% of the self-archived articles do not comply with publisher policy. In Chemical Engineering (CE) 77% of the self-archived articles do not comply with publisher policy. In Civil and Environmental Engineering (CEE) 83% of the self-archived articles do not comply with publisher policy.

Recall that in the context of this study, compliance refers strictly to whether the publisher policy allows or prohibits self-archiving and whether the policy requires, allows or prohibits self-archiving the publisher version of the work. Given the many other conditions and restrictions in publisher policies, the assessment of non-compliance in this study is no doubt lower than the reality of non-compliance.

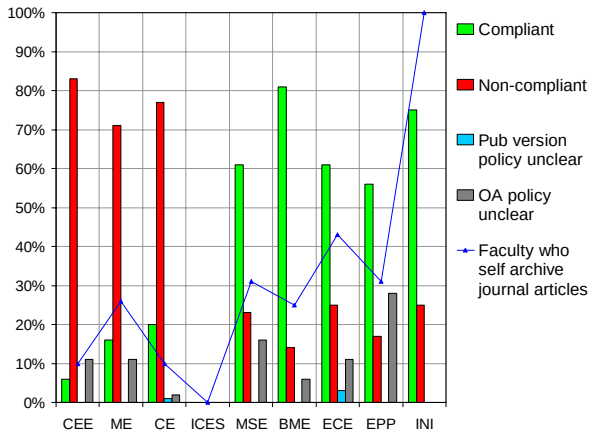


Figure 4: Compliance with publisher policy in CIT.

In Civil and Environmental Engineering (CEE), 93% of the non-compliant self-archived articles were published by publishers who prohibit self-archiving. In Chemical Engineering (CE), 55% of non-compliant self-archived articles were published by publishers who prohibit self-archiving. In Mechanical Engineering (ME), 32% of the non-compliant self-archived articles were published by publishers who prohibit self-archiving. In Materials Science and Engineering (MSE), 7% of the non-compliant self-archived articles were published by publishers who prohibit self-archiving. All other instances of non-compliance were related to the version that was self-archived. Either the publisher prohibited self-archiving the publisher version and the faculty self-archived the publisher version, or the publisher required self-archiving the publisher version and the faculty self-archived the author version. See Figure 5.

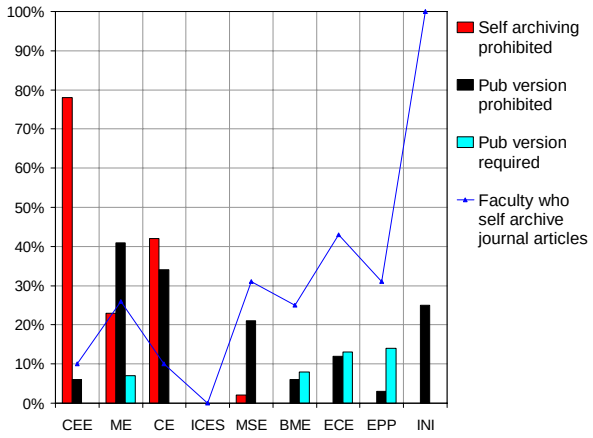


Figure 5: Analysis of non-compliance with publisher policies in CIT.

Figure 6 shows an analysis of the policies located for journals in which the engineering faculty publishes. Most publisher policies that allow self-archiving *prohibit* self-archiving the publisher version of the work. Some policies *allow* self-archiving the publisher version. A few policies *require* self-archiving the publisher version. In practice, most engineering faculty self-archive the publisher version of the work. The exception is Engineering and Public Policy (EPP), where self-archiving the author version seems to be more common. Many of the journals in which the engineering faculty publishes are published by the Institute of Electrical and Electronics Engineers (IEEE), which *requires* the publisher version to be self-archived.

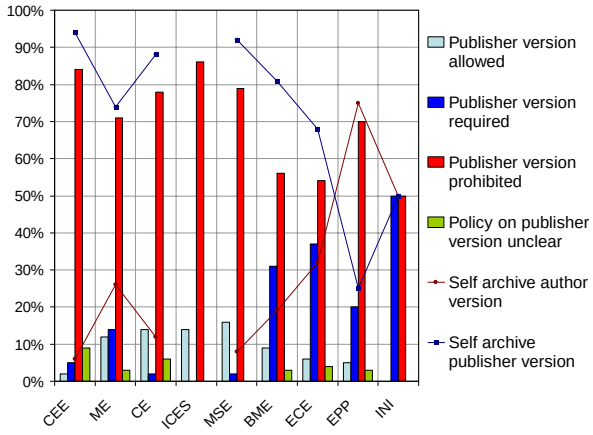


Figure 6: Analysis of journal self-archiving version policies and faculty practice in engineering.

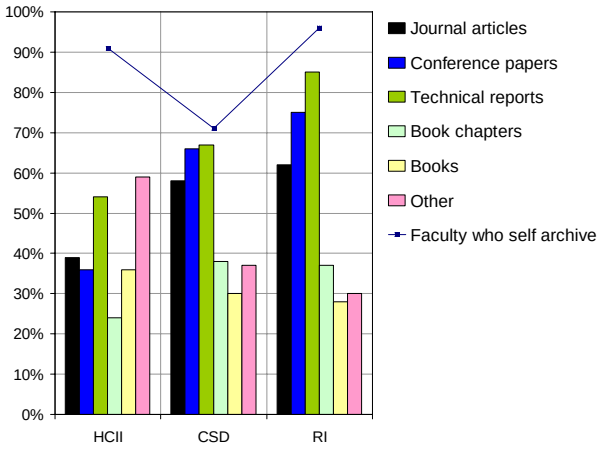
School of Computer Science (SCS)

To date, only three of the seven departments and research institutes in the School of Computer Science (SCS) have been examined: the Computer Science Department, the Human-Computer Interaction Institute and the Robotics Institute. These groups, however, account for 67% of the 206 faculty in SCS.

Based on the lists of publications found linked to the department web site, faculty in the Computer Science Department (CSD) and the Human-Computer Interaction Institute (HCII)

publish almost three times as many conference papers as journal articles and roughly twice as many journal articles as technical reports. The faculty in the Robotics Institute (RI) publishes almost four times as many conference papers and almost as many technical reports as journal articles. All three groups publish few book chapters, books or other materials. Overall, technical reports are the most likely to be self-archived.

Faculty behavior differs across the units. Figure 7 shows the self-archiving practice of faculty in HCII, CSD and RI. Left to right in the Figure, the data per department are organized by increasing self-archiving of journal articles. A larger percentage of faculty in HCII and RI (the research institutes) self-archives their work than faculty in CSD. Based on the SCS units examined to date, a larger percentage of computer science faculty self-archive their work than engineering faculty and, with rare exception, computer science faculty self-archive a larger percentage of their publications than engineering faculty. The notable exception is that the faculty in Materials Science and Engineering (MSE) self-archives a larger percentage of their journal articles than faculty in the Human-Computer Interaction Institute (HCII).



	DEPARTMENT OR INSTITUTE	FACULTY
HCII	Human-Computer Interaction Institute	22
CSD	Computer Science Department	63
RI	Robotics Institute	53

Figure 7: Self-archiving practice in the School of Computer Science (SCS).

Figure 8 provides data about publisher self-archiving policies of the journals in which CSD, HCII and RI faculty publish. For purposes of comparison, the data on the departments and research institutes in the school of engineering are included in the graph. Left to right in the Figure, the data are organized by increasing opportunity to self-archive journal articles based on publisher policy within the two schools. Data for the three departments and research institutes in SCS appear on the far right. Overall, more publishers of computer science journals allow self-archiving than publishers of engineering journals. Very few of the journal publisher policies in computer science prohibit self-archiving. No policy located prohibited self-archiving in the Human Computer Interaction Institute (HCII).

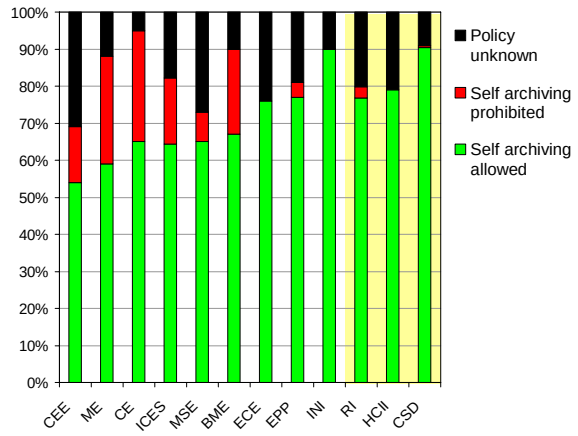


Figure 8: Journal publisher self-archiving policies in computer science compared with the policies in engineering.

Figure 9 provides details about the opportunity to and practice of self-archiving in the SCS units examined to date. The Figure also shows the percentage of faculty who self-archive journal articles in comparison with the percentage of faculty who self-archive any type of publication. Again, for purposes of comparison, the data from the school of engineering are included in the graph. Data for the SCS departments and research institutes appear on the far right.

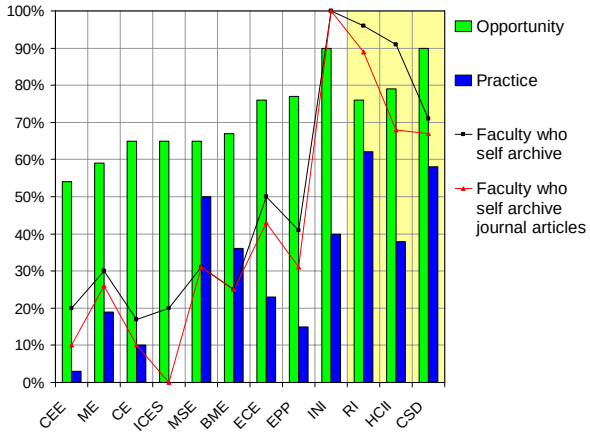


Figure 9: Opportunity to self-archive and practice of self-archiving journal articles in computer science compared with engineering.

Like faculty in most of the units in the school of engineering, fewer faculty in computer science self-archive journal articles than self-archive other types of publications. Faculty in the Computer Science Department (CSD) self-archived 58% of the journal articles cited in lists of publications linked to the department web site; publisher policy would allow them to self-archive 90% of them. Faculty in the Human Computer Interaction Institute (HCII) self-archived only 38% of their journal articles; publisher policy would allow them to self-archive 79%. The faculty in the Robotics Institute (RI) appears to be reaching almost their full potential to self-archive journal articles, having self-archived 62% of them with publisher policy allowing 76% of them to be self-archived.

As with faculty in the school of engineering, the data suggest that at least some faculty in the School of Computer Science are not aware of or do not care about publisher policies. See Figure 10. Roughly 28% to 32% of what is self-archived in RI, HCII and CSD does not comply with publisher policy.

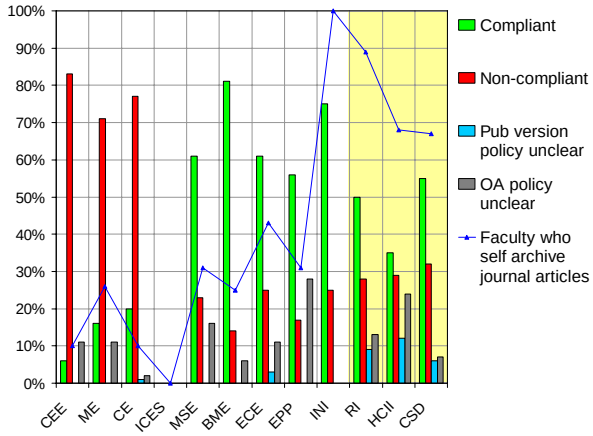


Figure 10: Compliance with publisher policies in computer science compared with compliance in engineering.

Figure 11 compares the type of non-compliance in the three computer science units examined to date with the type of non-compliance in the school of engineering. The percentage of publishers that allow self-archiving is much greater in computer science than in engineering. Recall that non-compliance in engineering was frequently the self-archiving of articles published by journal publishers who prohibit self-archiving. Few instances of non-compliance in SCS are because the publisher prohibited self-archiving. Most instances of non-compliance in SCS are because faculty self-archived the publisher version, which is prohibited by publisher policy. The remaining non-compliant articles are because the faculty self-archived the author version when the policy required self-archiving the publisher version.

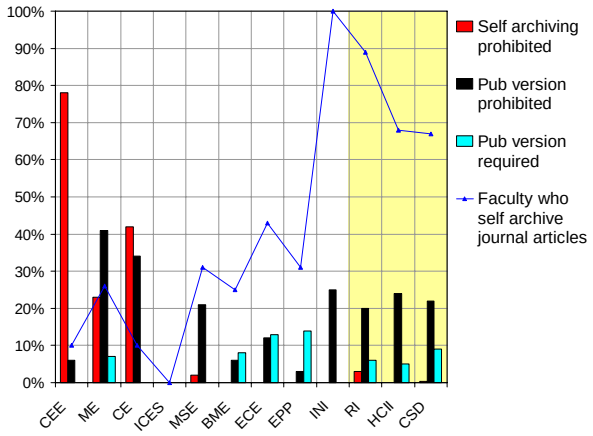


Figure 11: Analysis of non-compliance with publisher policies in computer science compared with non-compliance in engineering.

As noted earlier, most publisher policies that allow self-archiving *prohibit* self-archiving the publisher version of the work. A few *require* self-archiving the publisher version. In practice, the computer science faculty shows no strong preference for self-archiving the author or publisher version. See Figure 12. It may be the case that computer science faculty self-archive the author version more frequently than engineering faculty because they are more enthusiastic or habitual self-archivers and therefore perhaps more likely to self-archive earlier in the publication process, before the publisher version is available.

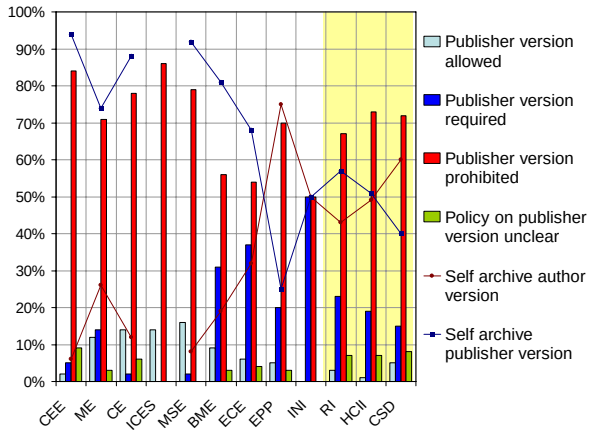


Figure 12: Analysis of journal self-archiving version policies and faculty practice in computer science compared with engineering.

Preliminary Conclusions

The data gathered and analyzed to date suggest that faculty either do or do not self-archive, regardless of publisher policy. There are likely multiple factors influencing their behavior, ranging from peer pressure and ideological positions for or against open access to issues of skill, tools and time. Carnegie Mellon University Libraries is currently conducting a small interview study to shed light on why faculty members do not self-archive their work.

After a faculty member has made the decision to self-archive, one can only speculate about how they choose what to self-archive. Factors that could influence decisions about what to self-archive include self-selecting:

- Their best work [2]
- Their most recent work
- Work they have in electronic format
- Work they have in any format (In the current study some faculty had scanned and self-archived work done years, even decades ago.)
- Work they have the time or assistance to self-archive

Looking specifically at the self-archiving of journal articles, the opportunity to self-archive in compliance with publisher policy and the practice of self-archiving varies significantly across disciplines. There appears to be no direct correlation between the opportunity to self-archive and faculty practice of self-archiving. Many faculty members appear not to know or not to care

about publisher policies prohibiting the practice of self-archiving or mandating what version of a journal article can be self-archived. Faculty members who self-archive do not consistently self-archive articles they publish in the same journal or self-archive the same version of articles they publish in the same or different journals.

Though not carefully analyzed, in doing Phase II of the study I observed that few faculty members attend to the subtle nuances of publisher self-archiving policies regarding required text to be displayed before or after publication of a work, the removal of pre-prints after articles are published, links to be provided to the publisher's website, etc. The evidence suggests that faculty update their publication lists on the web and self-archive their work intermittently. They do not, for example, routinely go back and change descriptive text or replace article pre-prints or early technical report versions with article post-prints or publisher PDF files. Busy faculty probably do not create and follow a schedule of publication dates and embargo periods to guide their self-archiving practice in compliance with publisher policy. Variations and picayune conditions and restrictions in publisher policies are likely too complicated to encourage or secure total compliance from faculty. Full compliance with publisher policies would be a scheduling and maintenance burden. The burden probably contributes to non-compliance and the gap between opportunity and practice. One has to wonder whether the burden is a publisher strategy to discourage self-archiving in practice while giving the impression of supporting open access. The variations and picayune details in publisher policies certainly make assessment of compliance a difficult task.

Next Steps

Phase I and II of the study will be completed for the remaining departments and research institutes in the School of Computer Science and all of the units in Mellon College of Science, College of Humanities and Social Sciences, Tepper School of Business, Heinz School of Public Policy, and College of Fine Arts. Data from the study will be used to inform faculty of the opportunity to self-archive in their discipline – focusing on the journals in which they publish – and to direct and fine-tune outreach and educational activities. The study results will help inform contributions and responses to policy proposals and other environmental developments, and to organize and prioritize faculty attention through formal and informal campus venues, including the Authors' Rights and Wrongs program. The University Libraries' goal is to work with faculty to remove barriers and to close the gap between the opportunity to self-archive and their practice of self-archiving.

Data from the study could be analyzed to better understand policy compliance per faculty member. The data provide a clear picture of how many different publisher policy variations the faculty need to know and the work that would be entailed to ensure full compliance. The data gathered in the study could also be analyzed by faculty track, rank on the track, and gender. This would enable a better understanding of trends, integration with findings from the 2006 faculty study of scholarly communication practices, and facilitate application of the findings. Future work might also include comparing the Phase II findings with the results from studies of the impact advantage of open access in different disciplines to determine whether faculty self-archiving practice is more frequent in disciplines where it is likely to be more effective in terms of their personal goals of promotion and tenure.[3]

References

[1] Findings from Carnegie Mellon University Libraries' 2006 study of faculty scholarly communication practices were not published. However, the results parallel those of a study conducted by the University of California. See *Faculty Attitudes and Behaviors Regarding Scholarly Communication: Survey Findings from the University of California*, prepared by UC Office of Scholarly Communication and the California Digital Library eScholarship Program in association with Greenhouse Associates, Inc., August 2007. Available at http://connect.educause.edu/library/abstract/FacultyAttitudesandB/45197.

[2] Research indicates that authors frequently self-select their higher quality work to be self-archived. See, for example, "The Self-Archiving Impact Advantage: Quality Advantage or Quality Bias?" by Stevan Harnad, available at http://eprints.ecs.soton.ac.uk/13193/01/kurtz-moed.html. See also, "OA Impact Advantage = EA + (AA) + (QB) + QA + (CA) + UA" by Stevan Harnad, available at http://eprints.ecs.soton.ac.uk/12085/01/OAA.html.

[3] See, for example, Yaşar Tonta, Yurdagül Ünal and Umut Al, "The Research Impact of Open Access Journal Articles," *Proceedings ELPUB 2007 Conference on Electronic Publishing*, Vienna, Austria, June 2007. Available at http://eprints.rclis.org/archive/00009619/.