
Measuring the Impact of Digitizing 24,000 Print Theses and Dissertations at UMass Amherst

Jessica Adamick
University of Massachusetts Amherst, USA

Abstract

The University of Massachusetts Amherst Libraries have begun to digitize and make open access our entire collection of 24,000 print theses and dissertations. Two years into the 10-year project, we can assess a variety of factors. There is a large body of data available for analysis of author responses, download counts, circulation counts, and donations. Low author opt-out rates demonstrate author support for the project, and high download counts demonstrate the immediate impact of an open access format for theses and dissertations. A successful development strategy is clearly identified. All of these data legitimate this massive digitization project.

Introduction

The entire collection of over 24,000 print theses and dissertations at the University of Massachusetts Amherst will be digitized and made open access over the next decade, having a huge effect on the visibility and use of these research products. Digitizing a thesis and dissertation collection is an enormous undertaking that requires human and financial resources; nearly every department in the UMass Amherst Libraries has been involved with this project.

Few universities have embarked on the digitization of print theses and dissertations. As more universities have adopted open access models for electronic theses and dissertations in their institutional repositories, it is only fitting that we work to increase the open access availability of the entirety of our theses and dissertations collections instead of having an accessibility gap between print and digital.

There are several components to this massive digitization project: selection of works to digitize and associated outreach to departments, contacting authors, development, processing materials, digitization, preparation for dissemination in software systems, dissemination through our

institutional repository, and follow-up with authors. There are opportunities to assess various aspects and outcomes of this project at nearly every stage. This paper examines the impact of this project by comparing print circulation and download statistics with a disciplinary analysis, publication date analysis, and digital version upload date analysis. It also reviews author donations and responses to project notification letters.

Having the data on hand that demonstrate increased use of these unique research products can effectively legitimate this work. Demonstrating author enthusiasm for open access is also beneficial for supporting this type of project. Early results from the project may help other institutions decide if digitizing their theses and dissertations would be worthwhile, based on use and author engagement.

Literature Review

Little information beyond a few key articles was found about the assessment of retrospective theses and dissertations projects. A study at the University of Arizona calculates the total cost of ownership of an electronic theses and dissertations project and includes a print versus digital version analysis, but the library is not working on digitizing print theses and dissertations, and so those are not part of the analysis.¹ A process and cost analysis of digitizing dissertations for an institutional repository at the University of Massachusetts Medical School includes an author permissions response analysis and an overall print circulation versus digital download analysis as well as the digitization cost calculations.² At the UMass Medical School, 74% of authors gave permission to post their dissertation open access, and downloads were 24 times the amount of print circulation in just 17 months versus five years. Finally, the London School of Economics performed a study exploring the impact of the online dissemination of theses on an institution's research profile, and the place of digital theses in scholarly communication.³ Working with a set of 2,000 theses,

the authors analyze permissions responses: 0.7% opted out of open access and five takedown notices were received. No correlation was found between the number of downloads a work received and the number of citations it received. The authors also analyzed traffic sources and conducted focus groups with undergraduates, postgraduates, and librarians on the awareness and perceptions of digital theses.

Project Overview

After a pilot program to test a basic workflow, the UMass Amherst Libraries' retrospective digitization project implemented the following process in earnest in 2014. Selected theses and dissertations are shipped to the Internet Archive's Open Content Alliance for digitization. We handle all other components of the project in-house.

Departments are selected for digitization each year, and we have the budget and capacity to digitize about 2,500 works annually. These works are uploaded to our institutional repository, ScholarWorks@UMass Amherst (ScholarWorks), where born-digital electronic theses and dissertations are also hosted. We select works for digitization by department (instead of publication year, for example), due to the opportunities for outreach and development. Another criterion for selection is department size; we select a variety of departments whose works combined equal about 2,500. The following departments have been selected for digitization so far, representing a diverse set of disciplines: Afro-American studies, astronomy, Chinese, education, history, polymer science and engineering, philosophy, political science, and psychology. Once a department is selected, the deans and department heads are then notified of the project and that the libraries will be contacting their alumni.

Our copyright analysis allows the libraries to post all theses and dissertations in our institutional repository, because the libraries already distribute print theses and dissertations through circulation and interlibrary loan, and only the format of work has changed. However, authors are contacted, notified of the project, and given an opportunity to opt in or out of having their thesis or dissertation made open access. Their responses are recorded. If an author opts in, their work is digitized and made open access through our institutional repository, and they are sent a link to their work. If an author opts out, their work is still digitized for preservation

and lending purposes, but their work is only made available digitally through on-campus access or through interlibrary loan services.

In the notification letters, it is suggested that authors donate the average cost of digitization (\$50) to support the project. For the first year of the project, donations were suggested at the point when the author was sent the link of their digitized thesis or dissertation. This included Afro-American studies, astronomy, Chinese, history, and psychology. One exception was polymer science and engineering; because that department was celebrating a major anniversary, the libraries worked with the department to fundraise, and the polymer science and engineering department collected the donations themselves. For the second and third years of the project, authors were notified of the opportunity to donate at the point of first contact. This approach applies to education, political science, and philosophy.

Methodology

Over two years of the project have been completed and all available author response, download, circulation, and donation data have been analyzed. As of September 2016, about half of the author notification letters from the education, philosophy, and political science batch had been sent out.

A major challenge in analyzing the project's data is the lack of common unique identifiers between the three main systems used: ScholarWorks (our institutional repository), Aleph (our integrated library system), and a development database. To allow analysis at the item level, records from the various systems had to be matched up, which is done in Microsoft Excel. We created unique identifiers based on fragments of fields such as author name, dissertation or thesis title, or graduation year, and then were able to match most records. However, there were numerous records without a match due to variations in fields comprising the unique identifiers. It is beyond the scope of this paper to manually match these records, so there are some missing data.

Tracking Author Responses

The astronomy, Chinese, history, polymer science and engineering, and psychology alumni have received notification letters and their responses are fully tracked and available for analysis. At the time of publication, about half of the education, philosophy, and political science alumni had

received notification letters, and responses from those alumni are available for analysis. Author responses are recorded in a “master file” in Excel that includes bibliographic data, author contact information, workflow steps, and, eventually, links to the institutional repository.

Tracking Download Counts

There are 2,012 works in astronomy, Chinese, history, polymer science and engineering, and psychology available for analysis by download count. Download counts are recorded through ScholarWorks, the institution’s institutional repository, which is run on bepress Digital Commons software. Metadata record hits are also available through the software, but this paper focuses on downloads of the actual work, since that measure is the closest analog to print circulation counts. Download counts are exported to Excel.

Tracking Circulation Statistics

There are 1,898 works available for circulation analysis. Circulation of print theses and dissertations is recorded through Aleph, the libraries’ integrated library system, and those data begin in 2006, the year that the libraries migrated software. These data are exported to Excel.

Tracking Author Donations

Ninety-nine donations were received and are available for analysis by donation rate. Donations are recorded in a development database and are exported to Excel.

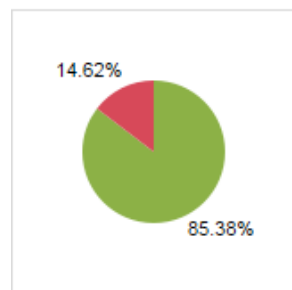
Findings

Author Responses

The great majority of authors (85.38%) opted in to the open access component of the digitization project (Figure 1).

Figure 1: Author Opt-in and Opt-out Rate

Author Opt-in and Opt-out Rate



Author response

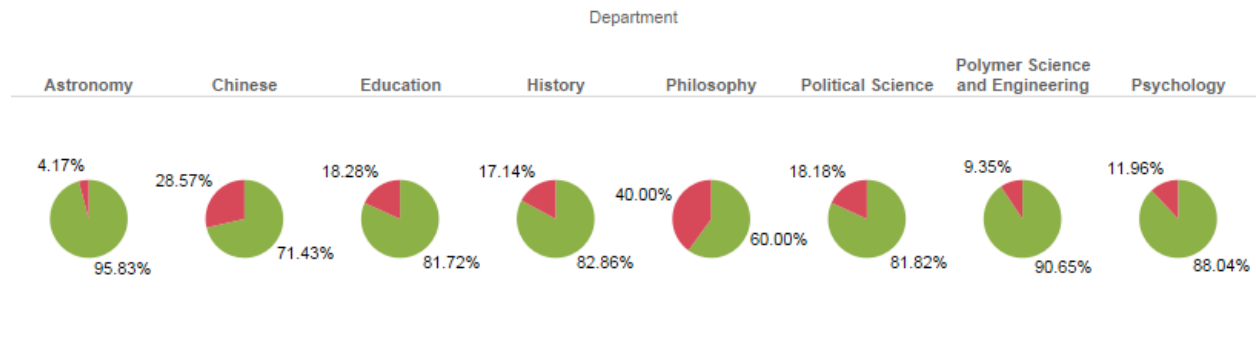
- Opt-in
- Opt-out

Across all departments, the majority of authors who responded to the notification letter opted in to the project (Figure 2). The departments with the highest opt-in rates are astronomy (95.83%), polymer science

and engineering (90.65%), psychology (88.04%), and history (82.66%). Philosophy has the highest opt-out rate at 40%, but not all authors had received their notification letters at the time of publication.

Figure 2: Author Opt-in and Opt-out Rate by Department

Author Opt-in and Opt-out Rate by Department



Author response

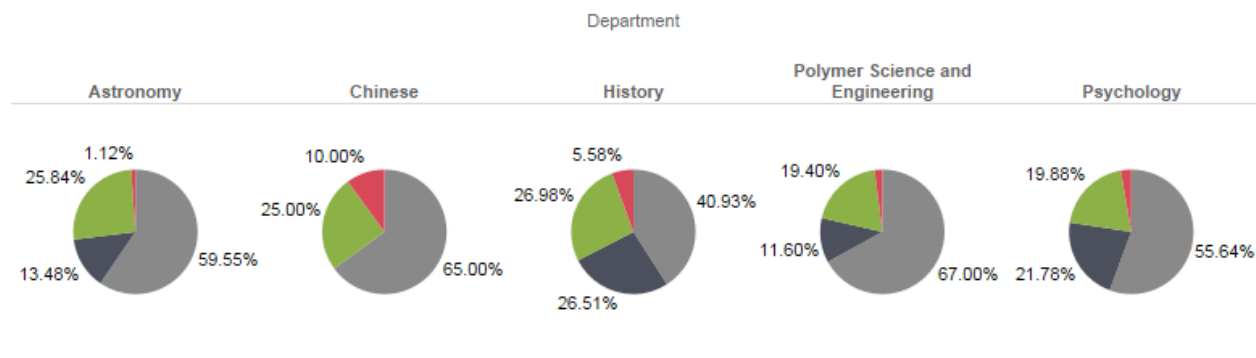
- Opt-in
- Opt-out

While we had author contact information for most authors in every department, across disciplines, most authors (56.38%) who received a digitization notification letter did not respond (Figure 3).

Education, political science, and philosophy are excluded from this figure, since not all alumni had received their letter at the time of publication.

Figure 3: Author Responses by Department

Author Responses by Department



Author response

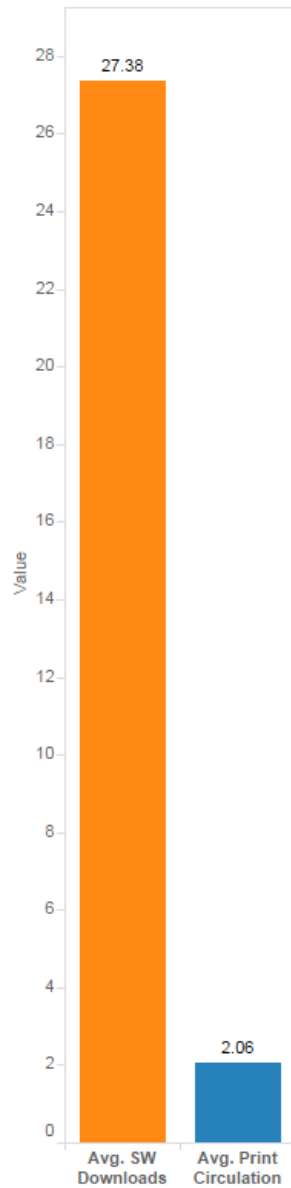
- No response
- No address
- Opt-in
- Opt-out

Download and Circulation Counts

The average download count for digitized theses and dissertations in ScholarWorks is 27.38 downloads, and the print circulation of the digitized theses and

dissertations, on average, is 2.06 times (Figure 4). Institutional repository download data is from the previous two years, and circulation data is from the last 10 years.

Figure 4: Average Download Count vs. Print Circulation
Average Download Count
vs. Print Circulation



Measure Names

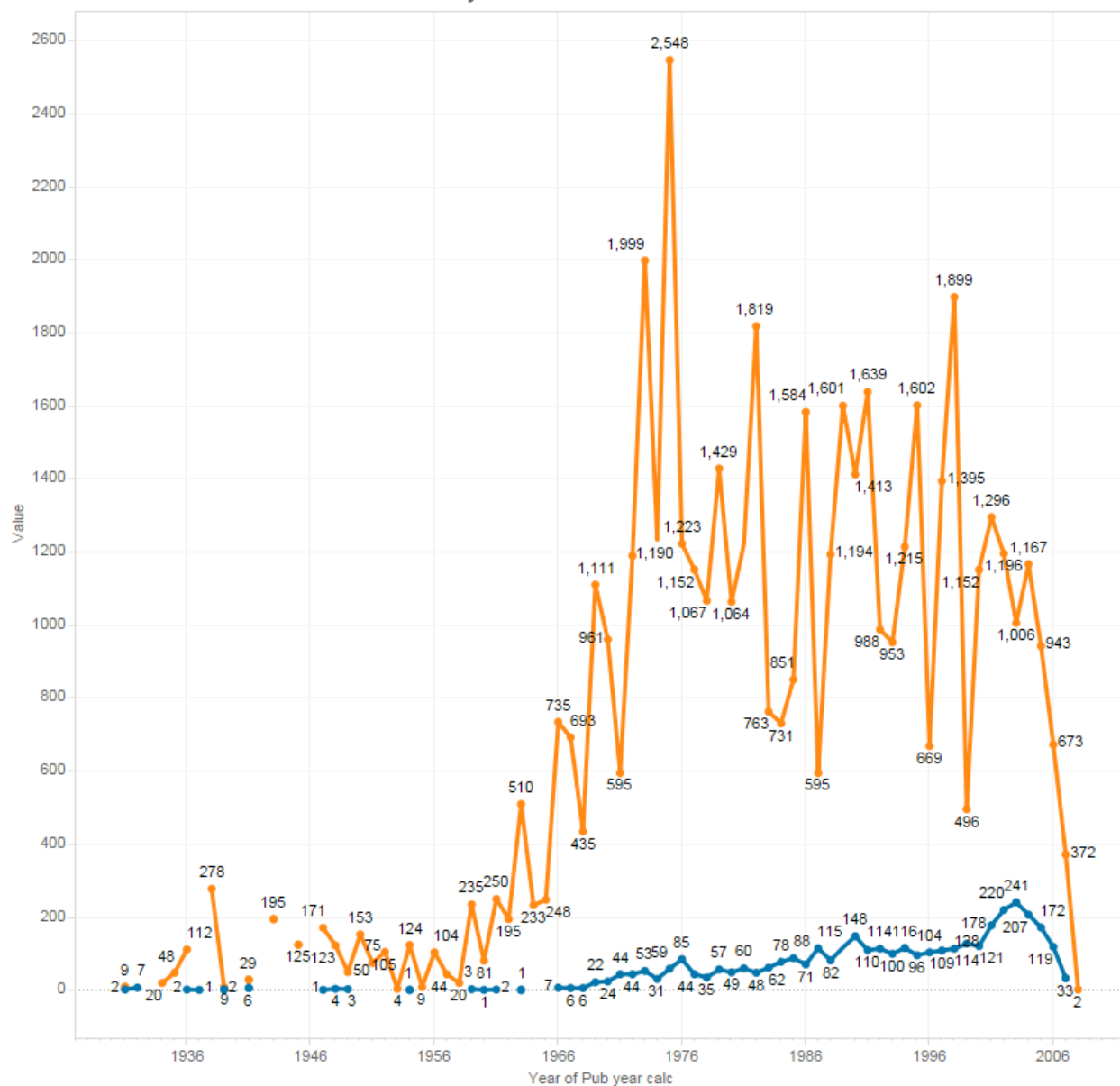
- Avg. SW Downloads
- Avg. Print Circulation

Recently published theses and dissertations receive more total download counts and total circulations than older theses and dissertations (those published

prior to the 1970s) (Figure 5). Still, the older works are seeing more digital use than print use.

Figure 5: Total Download Count versus Print Circulation by Publication Year

Total Download Count vs. Print Circulation by Publication Year



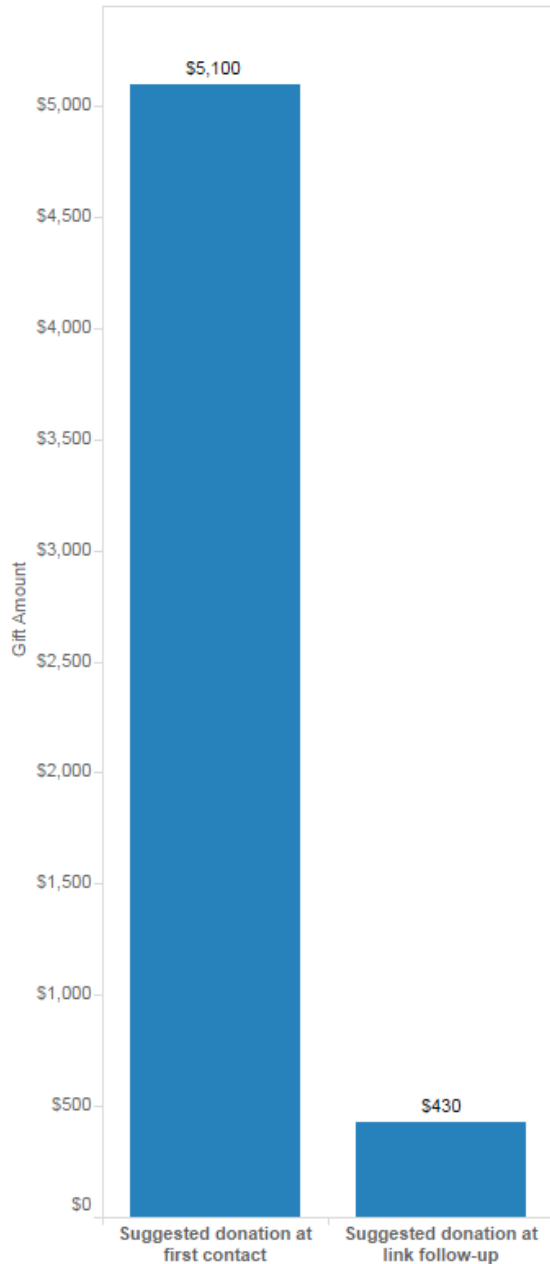
Measure Names
■ Print Circulation
■ SW Downloads

Author Donations

Two strategies were used to collect donations to support the digitization project (Figure 6). The first year, authors were sent a notification letter, and we sent the authors who opted in a link to their digitized work when it was complete. We suggested

a donation at the time they received the second contact. This strategy produced \$430 in donations. The second strategy involved suggesting a donation with the point of first contact with the digitization notification letter, and this produced \$5,100 in donations.

Figure 6: Suggested Donation Models and Associated Donations
Suggested Donation Models and Associated Donations



Discussion

Author Responses

Every department's alumni responded to the digitization notification letter with significantly more opt-ins than opt-outs. Some of these responses make sense in the context of what we know about the disciplinary practice of sharing scholarship, like the very low opt-out rate for astronomy and

the slightly higher opt-out rates for humanities disciplines like history and Chinese.

The biggest takeaway related to author responses is that we can confidently proceed with the project without worrying that alumni will be upset if their works are digitized and made available in an open access manner. We considered at the start of the project that it would be possible that many alumni

would opt-out of open access, considering that many of them developed their thesis or dissertation in an entirely different scholarly communication context before the advent of the Internet.

Download and Circulation Counts

While it will be no surprise to the library community, digital downloads of theses and dissertations dwarf print circulation of these works. In just two years, the digitized theses and dissertations have accumulated, on average, 13 times the amount of use than the print copies received in 10 years. This immediate increase in use clearly illustrates the elevated profile that these valuable research products receive once they are available not just digitally, but in an open access format.

The older works (published prior to the 1970s) receive less use in both digital and print format, which may be interpreted as the more recent scholarship being part of active scholarly communication, whereas the older works may have transitioned to archival sources after a few decades. The older works have still seen increased use and they also benefit from digital preservation strategies.

Author Donations

Between the two models for collecting donations, it is clear that suggesting a donation to all letter recipients at the point of first contact generates a much higher positive response, and that is the strategy that will be used going forward.

Each area of analysis produced very positive and affirming results, which are beneficial in a number of ways. We can see direct support of the project through author responses; the high opt-in rate demonstrates an acceptance of open access across generations, since many alumni developed their thesis or dissertation before the advent of the Internet or open access. Anecdotally, individual author responses have been overwhelmingly positive; many alumni have contacted staff who work on the project to offer their gratitude for the digitization service the libraries are providing, and sometimes to reminisce about their time at UMass. The opportunities for alumni and departmental engagement have exceeded our expectations.

The use analysis demonstrates immediate impact; the majority of print theses and dissertations in our

collection have not circulated, while nearly every thesis or dissertation in ScholarWorks has been downloaded. The highest print use for an item in this batch of digitized works is 15 circulations, while the average download count is 27.38 downloads. While not surprising, the use of the online versions is affirming to the project.

Finally, this is a massive project that nearly every department in the library touches, and being able to share some positive early data with library staff is a great opportunity to demonstrate the benefits of this project to those who work on it.

Conclusion

Analyzing the data produced by a massive theses and dissertations digitization project at the University of Massachusetts Amherst Libraries reveals very positive early results. Authors, when they did respond to the digitization notification letters, overwhelmingly opted in to the project, demonstrating an affinity for open access. Downloads in the past two years dwarfed 10 years of print circulation by a factor of 13 to one, demonstrating the value of digital and open access formats. Observing such high immediate use affirms the project beyond preservation goals. We were also able to identify a more successful development strategy of suggesting a donation upfront to all authors, which will be the strategy going forward.

—Copyright 2017 Jessica Adamick

References

1. Yan Han, "ETD: Total Cost of Ownership—Collecting, Archiving and Providing Access," *Library Management* 35, no.4/5 (2014), <http://dx.doi.org/10.1108/lm-08-2013-0084>.
2. Mary E. Piorun and Lisa A. Palmer, "Digitizing Dissertations for an Institutional Repository: A Process and Cost Analysis," *Library Publications and Presentations* 94 (2008), <http://dx.doi.org/10.3163/1536-5050.96.3.008>.
3. Linda Bennett and Dimity Flanagan, "Measuring the Impact of Digitized Theses: A Case Study from the London School of Economics," *Insights* 29, no. 2 (2016), <http://dx.doi.org/10.1629/uksg.300>.