JLSC

ISSN 2162-3309 | JLSC is published by the Iowa State University Digital Press | http://jlsc-pub.org

Volume 12, 1 (2024)

Longitudinal characteristics of bepress Digital Commons R1 and R2 institutional repositories

Jimmy Ghaphery, Abbey Childs & Karen Bjork

Ghaphery, J., Childs, A. & Bjork, K. (2024). Longitudinal characteristics of bepress Digital Commons R1 and R2 institutional repositories. *Journal of Librarianship and Scholarly Communication*, 12(1), eP17786. https://doi.org/10.31274/jlsc.17786

This article underwent fully anonymous peer review in accordance with JLSC's peer review policy.

ISSN 2162-3309 10.31274/jlsc.17786

RESEARCH ARTICLE

Longitudinal characteristics of bepress Digital Commons R1 and R2 institutional repositories

Jimmy Ghaphery
Virginia Commonwealth University
Abbey Childs
Virginia Commonwealth University
Karen Bjork
Virginia Commonwealth University

ABSTRACT

Introduction: This study presents longitudinal findings on the size, usage, and other characteristics of institutional repositories of Carnegie R1 and R2 institutions using the Digital Commons bepress (bepress) platform.
Methods: In 2017, 2019, 2021, and 2023, a subset of bepress repositories (R1 and R2 institutions) were visited in order to gather data on number of items, downloads, journal presence, and theses and dissertations, as well as changes in the bepress customer base.
Results: During this period, the total number of R1 and R2 bepress institutions grew from 76 in 2017 to 90 in 2019, followed by declines to 86 in 2021 and 82 in 2023. From 2017 to 2023, 15 R1 or R2 institutions were observed leaving bepress, and 12 joined. These 84 institutions, on average, represented upward of 30% of all the Carnegie R1s and R2s and 17% of bepress customers.
Discussion: Results show a growing body of items within these repositories along with increased downloads. From 2017 to 2023, the bepress R1 and R2 institutions added a total of almost 1.8 million items and recorded more than 416 million downloads. The total number of items in the R1 and R2 bepress repositories was 3,061,447 in 2023. While the R1 institutions generally had more items, downloads per item were similar between R1 and R2 institutions. Theses and dissertations were downloaded at a higher rate than the rest of the items in the R1 and R2 bepress repositories. The presence of journals was found in more than 84% of the repositories, and almost all of the R1 bepress repositories.
Conclusion: The results paint a picture of potential repository needs with regard to technical scale and future growth needs, as well as provide some benchmarking data for individual repositories. The results also point toward repository needs in terms of features to support, manage, and potentially grow these collections. The observable capacity, impact, and needs of the bepress R1 and R2 repositories from 2017 to 2023 are especially

Received: 03/19/2024 Accepted: 06/03/2024

Conflict of Interest Statement: The authors work at an R1 institution library that uses bepress Digital Commons for its institutional repository. Neither bepress nor its parent company, Elsevier, has contributed to or reviewed this study.

relevant at a time when institutional repositories are maturing alongside ongoing conversations regarding ownership of scholarly communications infrastructure.

IMPLICATIONS FOR PRACTICE

- Outlines scale and impact of a large cohort of institutional repositories with implications for the scale needed for shared repository infrastructure
- Contributes to discussions on repository platform choice and use
- Provides benchmarking comparisons for local efforts as well as considerations for the evolution and priority of various aspects of repositories for the future

INTRODUCTION

Institutional repositories (IRs) serve as a core system for scholarly communications efforts. According to OpenDOAR, as of January 2024, there were more than 5,874 open access (OA) repositories around the world, and 7% use the Digital Commons platform (Jisc, n.d. [a]). This study looks at the scale and usage trends of a group of IRs on the bepress Digital Commons (bepress) system between 2017 and 2023. The selected group of, on average, 84 repositories represents an appreciable percentage of United States doctoral institutions with significant research activity. As such, the findings not only describe the specific group of bepress repositories but also provide some insight into the larger repository platform ecosystem with regard to capacity, features, impact, and migrations.

A longitudinal examination of IRs can help direct the investment that goes into choosing, maintaining, and optimizing these platforms. For example, statistics such as the number of items in the repository, item uploads, participating academic units, and contributing faculty help tell a story about the scale and distribution of scholarship at an institution. IR metrics can likewise inform potential IR requirements for shared open infrastructure, which have been of growing interest to the scholarly communications community since the Elsevier acquisition of bepress in 2017. IR metrics (including those mentioned above as well as the number of downloads and locations of visitors) can also be used to lobby for the allocation of funding and support for open scholarly infrastructure, potentially helping to overcome the challenge of convincing university administrators to support open scholarship. In addition to individual institutions' statistics, looking at data and trends across the open scholarship community on the usage of IRs demonstrates the scale needed for shared repository infrastructure and offers possible ways of measuring a return on investment for support for open infrastructure. This study seeks to provide insight into the current use of IRs as the profession engages with questions about the broader open scholarly communications infrastructure.

LITERATURE REVIEW

IR metrics

Some of the earliest work regarding IR measurement includes various surveys and studies, especially focused on how OA repositories might supplant traditional closed publishing practices (Bailey et al., 2006; Lynch & Lippincott, 2005; McDowell, 2007). Dubinsky's (2014) mixed methods study on IR use surveys more than 200 bepress repositories with quantitative measurements of growth over a six-month period as well as includes a survey of the repository administrators that was focused on faculty adoption. Dubinsky notes wide variation in growth rate among repositories with a median of two items added per day. Bruns and Inefuku (2016) detail various metrics by which IRs gauge engagement in scholarship at an institution, and the varied audiences and use cases of those measures.

The Association of College and Research Libraries has been collecting two IR metrics—items added and downloads—as part of their annual statistics collection since 2015 (Petrowski, 2015) which are available through subscription to the *Benchmark: Library metrics and trends* tool (http://www.ala.org/acrl/proftools/benchmark). There are also multi-institutional efforts to measure IR use. Institutional Repository Usage Statistics (IRUS) is based on the COUNTER Code of Practice and allows participating institutions to compare repository use (Jisc, n.d. [b]). IRUS started in the UK as the Publisher and Institution Repository Usage Statistics (PIRUS2) project in 2009. IRUS has grown to include 212 repositories, including repositories from the United States and Australia. The Repositories Analytics and Metrics Portal (RAMP), started in early 2017, is another effort to collect comparable statistics from various repositories. RAMP is based on the harvesting of Google Analytics data and currently contains data from more than 60 repositories (RAMP, n.d.). Demonstrating how RAMP could be used as an open IR usage dataset, Arlitsch et al. (2021) analyze the use and performance data of 35 IRs, including institutions from seven different countries and a wide range of IR sizes. The study finds variation in the performance and use of IRs, including some geographically bound differences and, relevant to this study's findings, that repositories with more theses and dissertations tended to have higher usage than those that contained fewer or none (Arlitsch et al., 2021).

IR ownership

A major facet of the discussion of open scholarly infrastructure is the issue of commercial versus shared ownership. Elsevier's 2017 acquisition of bepress raised major concerns about ownership within the OA landscape and left many librarians feeling betrayed by the acquisition. Barbara Fister explains: "One of the reasons so many librarians felt angry about Elsevier's acquisition of bepress is that Elsevier has become an enemy of libraries and what they stand

for" (Fister, 2017, para.1). The considerable discussions within the scholarly communications and library community led to calls for community-led efforts and proposals of shared responsibility and governance models. David Lewis' proposal for a shared financial commitment states, "Every academic library should commit to contribute 2.5% of its total budget to support the common infrastructure needed to create the open scholarly commons" (Lewis, 2017, p. 1). Lewis outlines the role of the academic library in the broader open scholarly commons, the potential impact of meaningfully funding the development of infrastructure to support the commons, and a plan for implementation that includes possible stakeholder organizations and ways to incentivize participation.

Neylon (2018) offers a critique of Lewis' proposal, noting that Lewis' model is both "too ambitious and not ambitious enough" and arguing that it is unlikely that enough universities will participate in the model. Neylon cites political risk, lack of sway over budgets, and lack of large-scale buy-in from the broader campus community as major hurdles to Lewis' plan and concludes that the 2.5% figure will be ultimately too small an investment to build adequate infrastructure to fit the needs of the scholarly communications community. Another notable example from the same time period, the Institute of Museum and Library Services funded project "It Takes a Village" looks closely at how cultural and scientific heritage organizations can sustain open source programs with attention to governance, technology, human resources, financial support, and community engagement (Lyrasis, n.d.). In discussing the results from a 2019 census of more than 60 scholarly communications tools and services, Skinner (2019) finds numerous sustainability challenges for infrastructure providers, including financial, governance, and vision.

Dohe (2019) contributes an important critical dimension to these infrastructure discussions with reminders that these systems cannot be separated from the wider power dynamics that are often at odds with the stated goals of open practices. Tillman's (2019) humorous but all too real depiction of the librarian perspective demonstrates the unseen costs of switching repository platforms in search of a better solution. While beyond the scope of this literature review, it is important to note that there is an established literature on infrastructure that is broader than and predates scholarly communications systems. The work of Goudarzi and Dunks (2023), which is part of the Invest in Open Infrastructure (https://investinopen.org/) project, is an important touchstone that both explores this broader literature and widens the infrastructure question beyond the technology to include the full suite of services around any solution.

Discourse conceptualizing shared infrastructure inevitably leads to discussions about ownership of that infrastructure, which have been further fueled by subsequent ownership changes across the open scholarly landscape. Notable examples include Taylor & Francis's acquisition of F1000 Research in 2020 (Bell, 2020) and Wiley's acquisition of Knowledge Unlatched in 2021 (Knowledge Unlatched, 2021). In response to these changes, Community-led Open

Publication Infrastructures for Monographs (COPIM) released a statement calling for structures and safeguards to ensure that open scholarly infrastructure is community-owned and governed (COPIM, 2021). COPIM cites concerns that these acquisitions "reflect an ongoing consolidation of research infrastructure by major publishing corporations, and in particular the increasing attempts to monetize and, potentially, monopolize the infrastructures of open knowledge dissemination" (COPIM, 2021, para. 1).

Analysis of IR use and performance data is still a relatively new topic of study but has implications for the discussion about open scholarly infrastructure. This paper aims to contribute to this ongoing conversation by documenting repository size and downloads among bepress R1 and R2 institutions. This study also provides insights into potential differences of R1 and R2 IRs and bepress market share among R1 and R2s, and it notes what platforms some R1 and R2 bepress institutions have migrated to during the time period of the study.

METHODS

The genesis of the study in 2017 followed the Elsevier acquisition in an attempt to gain some understanding of the characteristics of bepress research repositories. The consistent public display of repository statistics encouraged the authors to continue the study with this set of bepress research institutions in a longitudinal manner with biennial data collection.

Data was recorded into spreadsheets by visiting a selected group of bepress Digital Commons repositories in the fall of 2017, 2019, 2021, and 2023. The lead author gathered data in 2017, 2019, and 2021. All authors of this paper gathered the 2023 data. Using the Digital Commons public site, the authors copied a list of all existing bepress customer repositories and a subset of customers listed in the "Research" institution section of the bepress site along with the URL to each repository (bepress, n.d.). For the "Research" institutions, the following information was collected from each site:

1. Item and download data: Number of items, total cumulative downloads, total downloads in the past year
2. Item and download data for theses and dissertations: Number of items, total cumulative downloads, total downloads in the past year
3. Presence of journals as indicated by journal title and articles arranged in volume and issue format
4. Whether institutions were new to the set or had left since the last observation. For those institutions that had left, it was noted what year they had last been observed, and what if any repository system they had migrated to. Determination of the new platform was based on observation of the new interface and viewing html source code.

Item and download data were copied from the public webpages of each repository, both at the repository level and, where available, for subpages of theses and dissertations. These numbers (total items, total downloads, downloads in the past year) were copied from the bottom of part of the bepress readership map for each repository (see Figure 1).

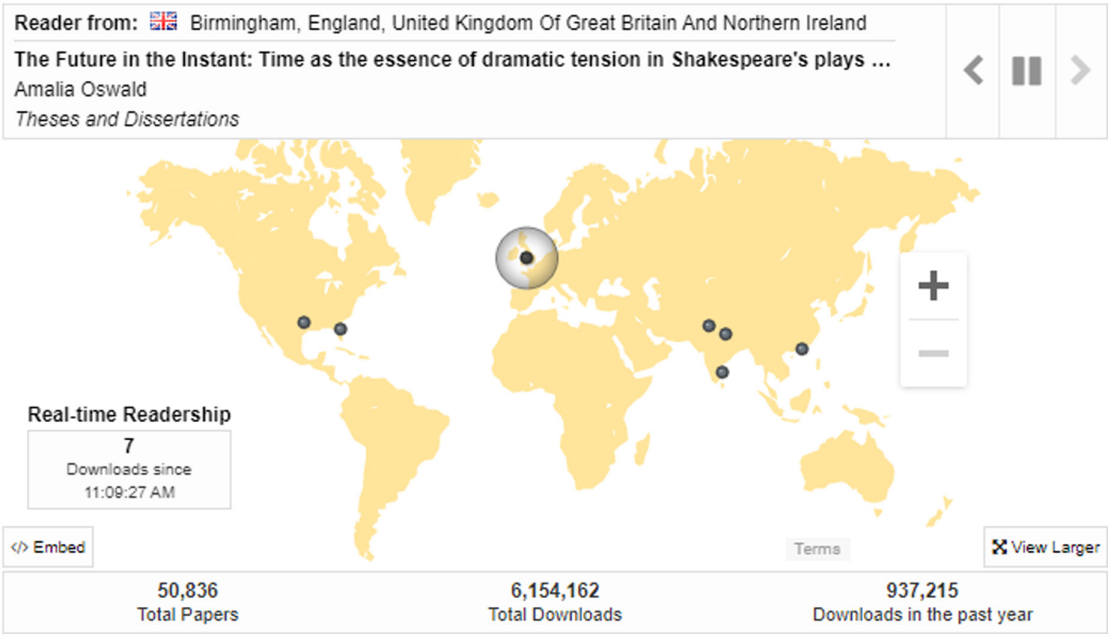

Figure 1. bepress readership map.

The Carnegie Classification of Institutions of Higher Education basic designation was added for each institution through lookups and cross-tabulation with the publicly available Carnegie downloads. For this study, only those institutions with the following basic Carnegie designations were included: "Doctoral Universities: Very High Research Activity" (R1); "Doctoral Universities: High Research Activity" (R2). For the 2017 dataset, the Carnegie basic classification at the time was "Doctoral Universities: Highest Research Activity" and "Doctoral Universities: Higher Research Activity." The list of institutions was further refined by eliminating professional schools and, in one case, a university that was using bepress exclusively for journal publishing and not other repository functions.

Using pivot tables and other spreadsheet tools, data was compiled for each of the four years (2017, 2019, 2021, 2023) in order to answer the following research questions about these repositories:

- What percentage of bepress customers are Carnegie R1 and R2 institutions, how does that compare to the total number of Carnegie R1 and R2 institutions, and how has that percentage changed over the period of six years?
- How has the item count and download activity for these repositories changed over the period of six years?
- What percentage of items and downloads do theses and dissertations account for?
- What is the prevalence of journals?
- What differences present themselves between R1 and R2 repositories?

Limitations of the study include replication, as it would be difficult to reharvest institutional item and download data for previous years. The data recorded on the presence of theses and dissertations and journal publishing relied on the judgment of the researchers and what information was readily available on the IR site. However, the basic institutional demographic methods could be replicated, as lists of bepress Digital Commons institutions from the years studied are available through the Internet Archive, and historical Carnegie basic classification data is publicly available. The findings could be reproduced from the source data that is openly available through the Open Science Framework (Ghaphery et al., 2024).

Another methodological limitation is the reliance on the item and download data being available via bepress's readership map feature on the front page of the repository and for theses and dissertations. In all of the years studied, there was never an instance in which an institution had chosen to hide this information on the main institutional page. There were, however, instances in which this data was not available for theses and dissertations. While the item and download data is a like-for-like comparison of different institutions on the same platform, a limitation is that the data itself is provided and documented by the platform vendor, bepress. For the number of items, the bepress readership map does not include supplemental files. The vendor-specific nature of the dataset also limits the scope of the study to R1 and R2 U.S. institutions on the bepress platform. With regard to this limitation, there should be caution in generalizing to all repositories or even all U.S. repositories with very high and high research activity.

RESULTS

bepress Digital Commons customer base

The basic demographic characteristics provide the number of institutions in the dataset and their percentage relationship to the number of other bepress customers for 2017, 2091, 2021,

and 2023. For the four years observed, the total number of R1 and R2 doctoral institutions on the bepress platform in comparison to all bepress customers was 17.17% on average: 76 of 445 total in 2017 (17.08%), 90 of 481 in 2019 (18.71%), 86 of 511 in 2021 (16.83%), and 82 of 511 in 2023 (16.05%). Note that while the total number of customers for the 2021 and 2023 observations was 511, the actual institutions differed, reflecting customers who had left or joined the platform since 2021 (see Figure 2).

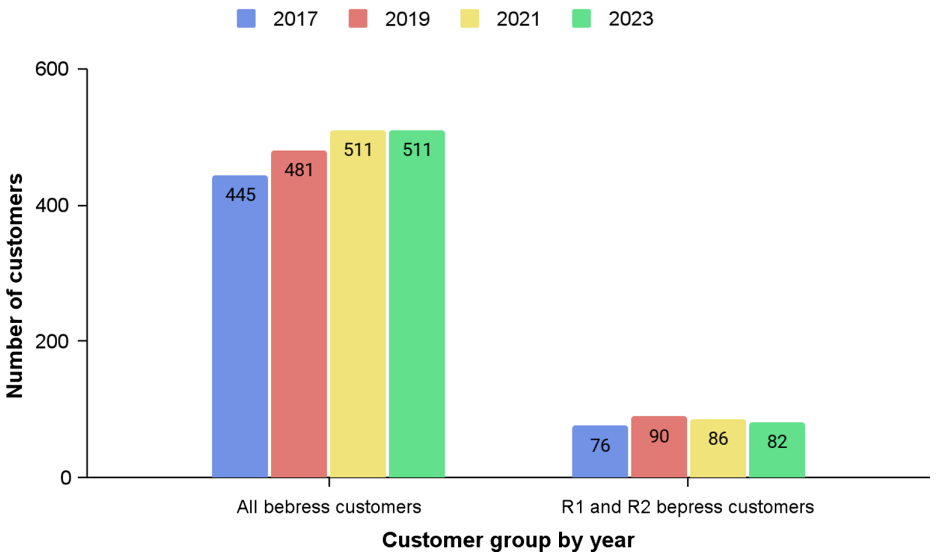

Figure 2. bepress Digital Commons customers.

Comparing the bepress R1 and R2 institutions with the complete national list of institutions with those Carnegie classifications demonstrates that, on average, across all four observable years, the bepress institutions account for 32.51% of R1 and R2 institutions (R1 28.89%; R2 36.22%). The bepress to Carnegie comparison by year for R1 and R2 institutions was 76 bepress R1/R2s of 222 total Carnegie R1/R2s in 2017 (34.23%), 90 of 265 in 2019 (33.96%), 86 of 265 in 2021 (32.45%), and 82 of 279 in 2023 (29.39%). Over the six-year period, the total number of bepress R1s increased from 33 to 41, and the number of R2s decreased from 43 to 41, while the total number of Carnegie R1 and R2 institutions increased (R1s from 115 to 146 and R2s from 107 to 133) (see Table 1). The growth of R1 and R2 bepress institutions in 2019 was not entirely attributable to new or leaving customers. There were nine new institutions in the dataset who were already bepress customers in 2017, but their Carnegie basic classification changed from "Doctoral Universities: Moderate Research Activity" to "Doctoral Universities: High Research Activity." There were four institutions whose Carnegie basic designation changed from "Doctoral Universities: Higher

Research Activity" (R2) in 2017 to "Doctoral Universities: Very High Research Activity" (R1) in 2019. Carnegie designations for the bepress R1 and R2 institutions in 2021 and 2023 were stable.

	Carnegie R1 Institutions		Carnegie R2 Institutions	
Year	R1 bepress Customers	All R1 Institutions	R2 bepress Customers	All R2 Institutions
2017	33	115	43	107
2019	39	131	51	134
2021	38	131	48	134
2023	41	146	41	133

Table 1. bepress R1 and R2 Customers Compared to All R1 and R2 Institutions By Year

The results document institutional migrations to or from the bepress platform during the time period of 2017 to 2023 (see Figure 3). During that time, there were a total of 12 new R1 and R2 bepress customers observed: 8 in 2019 (4 R1, 4 R2), 3 in 2021 (2 R1, 1 R2), and 1 in 2023 (R1). All of the new institutions from 2019 and 2021 were still in the dataset in the fall 2023 survey. From 2017 to 2023, there were a total of 15 R1 and R2 institutions that had left the bepress platform (8 R1 and 7 R2) as follows: 2 R1 and 1 R2 in 2019; 3 R1 and 4 R2 in 2021; and 3 R1 and 2 R2 in 2023.

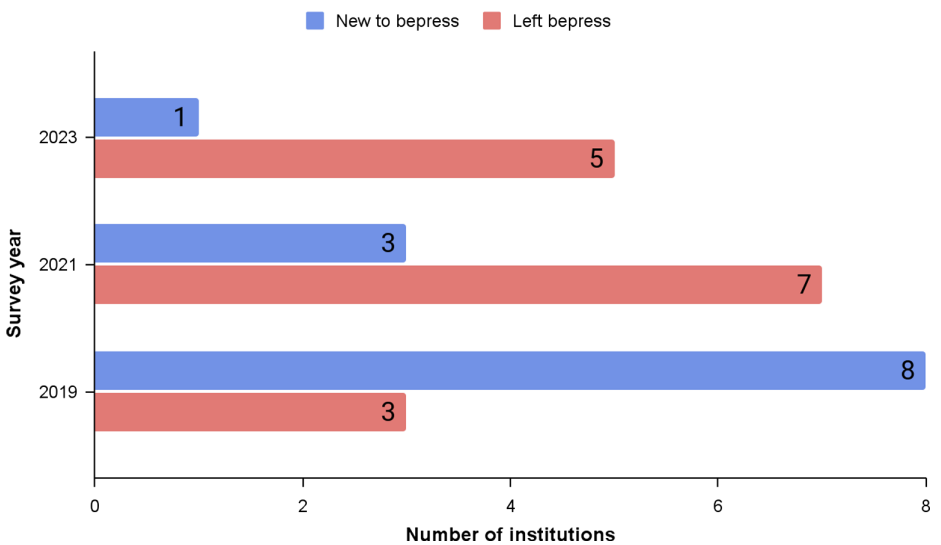

Figure 3. bepress repository migrations for R1 and R2 institutions.

New platforms for the 15 total R1 and R2 institutions that left bepress were DSpace (3 R1), Esploro (2 R1 and 1 R2), Islandora (4 R2), Samvera (2 R1 and 1 R2), Figshare (1 R1), and No repository (1 R2). Of the 14 institutions that moved to a new repository, 10 migrated to an open source solution (DSpace, Islandora, Samvera), with the remaining 4 opting for a commercial platform (Esploro, Figshare).

Repository size and usage

From 2017 to 2023, the bepress R1 and R2 institutions added a total of 1,794,252 items and recorded 416,424,183 downloads. The total number of items in the R1 and R2 bepress repositories combined was 1,267,195 (2017), 1,992,132 (2019), 2,555,365 (2021), and 3,061,447 (2023) (see Figure 4).

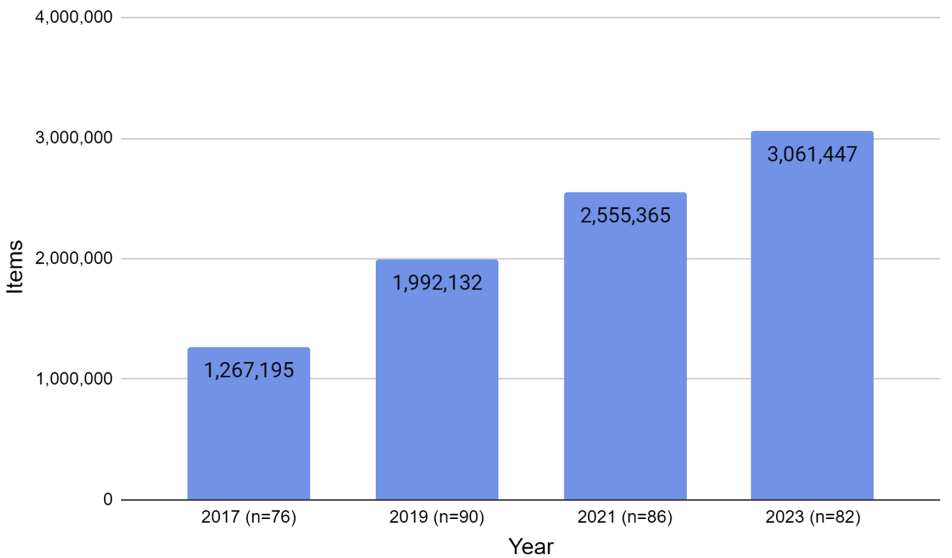

Figure 4. Total items in R1 and R2 bepress repositories.

There was a wide range of repository sizes, measured by the total number of items and total downloads, with the majority of repositories on the smaller side of average. Table 2 shows the average, median, and standard deviation values for total items in R1 and R2 bepress repositories by year.

Items	2017 (n = 76)	2019 (n = 90)	2021 (n = 86)	2023 (n = 82)
Average	16,674	22,135	29,714	37,335
Median	13,047	15,149	18,505	22,718
Standard Deviation	17,317	22,919	29,937	36,040

Table 2. Number of Repository Items R1 and R2 bepress Customers 2017–2023

The total number of yearly downloads observed for the bepress R1 and R2 repositories was roughly 53 million in 2017, 72 million in 2019, 107 million in 2021, and 88 million in 2023. Table 3 shows the average, median, and standard deviation values for total annual downloads in R1 and R2 bepress repositories by year.

Measures	2017 (n = 76)	2019 (n = 90)	2021 (n = 86)	2023 (n = 82)
Average	699,595	800,340	1,243,150	1,071,462
Median	391,106	505,489	905,817	851,235
Standard Deviation	1,036,617	1,141,163	1,530,297	1,010,168

Table 3. Total Annual Downloads (in the Past Year) for R1 and R2 bepress Repositories

Downloads in the past year per item were 41.96 (2017), 36.16 (2019), 41.84 (2021), and 28.70 (2023). Broken out by the R1 and R2 institutions for 2023, the R1 institutions accounted for 64% of the total repository items and 63% of the downloads in the past year (see Table 4).

	Items in repository		Item downloads	
Carnegie Classification	Number of items	Percent of total number of items	Number of downloads	Percent of total downloads
R1 (n = 41)	1,958,689	63.98%	55,303,493	62.95%
R2 (n = 41)	1,102,758	36.02%	32,556,402	37.05%
Total	3,061,447	100.00%	87,859,895	100.00%

Table 4. Total Item Number and Yearly Downloads, R1 and R2 bepress Repositories, 2023

Details such as average, median, and maximum values portray additional differences between R1 and R2 institutions for 2023 total items and downloads in the past year (see Table 5).

Item and Download Distribution	R1 (n = 41)	R2 (n = 41)
Total Items Average	47,773	26,897
Total Items Median	29,142	18,503
Total Items Maximum	161,839	99,513
Downloads in Past Year Average	1,348,866	794,059
Downloads in Past Year Median	1,112,088	706,723
Downloads in Past Year Maximum	7,157,816	2,666,098

Table 5. Distribution of Total Item Count and Download Count for bepress R1 and R2 Institutions, 2023

Theses and dissertations

On average, 69.81% of the bepress R1 and R2 institutions had observable item and download data for theses and dissertations: 53 of 76 in 2017 (69.74%), 61 of 90 in 2019 (67.78%), 60 of 86 in 2021 (69.77%), and 59 of 82 in 2023 (71.95%). Cumulatively from 2017 to 2023, this subgroup of bepress R1 and R2 institutions added 193,322 theses and dissertations and recorded 126,558,410 downloads for theses and dissertations. The total number of theses and dissertations across this group by year was 202,088 (2017), 280,197 (2019), 358,998 (2021), and 395,410 (2023). Theses and dissertations downloads in the past year were 15,667,955 (2017), 22,620,501 (2019), 36,480,398 (2021), and 23,382,054 (2023), with a yearly download per item count of 77.53 (2017), 80.73 (80.73), 101.62 (2021), and 59.13 (2023). In comparison with the total item count and total yearly downloads for the same subset of bepress R1 and R2 repositories with observable theses and dissertations data, theses and

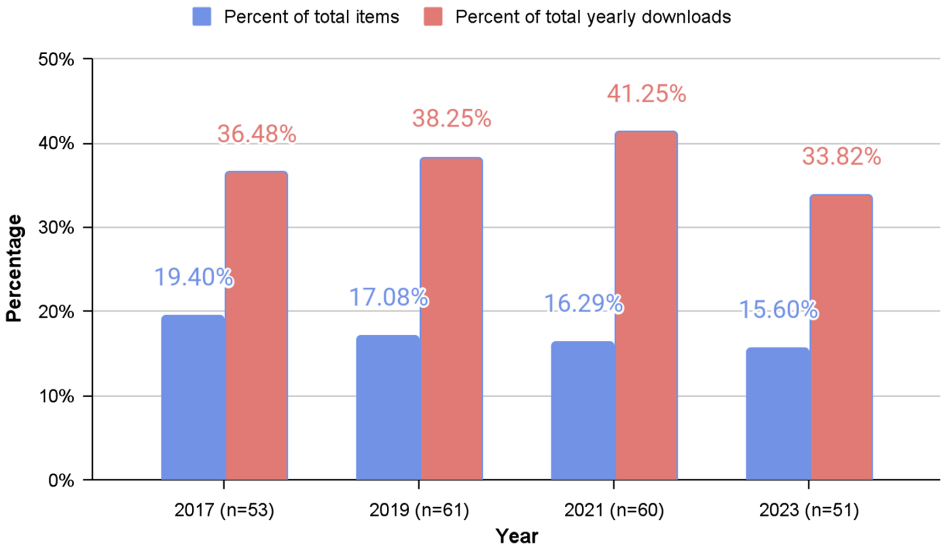

Figure 5. Theses and dissertations in bepress R1 and R2 repositories.

dissertations consistently accounted for at least 15% of the repository items and more than 33% of the total repository downloads (see Figure 5).

Journals

In visiting every R1 and R2 bepress repository, it was also noted whether or not the site was presenting journals using the bepress publishing platform whereby journal articles are organized under journal titles with volumes and issues. Figure 6 shows the percentage of repositories using bepress to host journals for R1 and R2 sites combined. For the R1 institutions, journal presence was found in 32 of 33 institutions in 2017 (96.97%), 37 of 39 in 2019 (94.87%), 37 of 38 in 2021 (97.37%), and 40 of 41 in 2023 (97.56%).

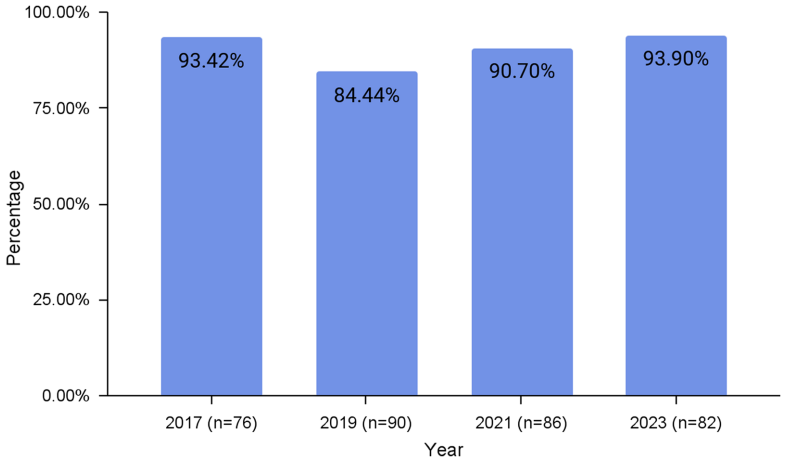

Figure 6. Percentage of R1 and R2 bepress repositories hosting journals.

DISCUSSION

It is not a coincidence that the study began in 2017 with the announcement that Elsevier had purchased bepress. As noted in the literature review, news of the acquisition accelerated considerable professional discussion about repository platforms and shared infrastructure. Six years later, across four observations, the bepress customer base appears to be stable.

For the total number of bepress customers, there was some growth from 2017 through 2021 and, for R1 and R2s, a combined growth from 2017 to 2019 and then a slight decline. In both cases (all customers and the R1/R2 group), the customer base was larger in 2023 than in 2017, with all customers growing from 445 to 511 (14.83% growth) and R1/R2 customers growing

from 76 to 82 (7.89% growth). This finding, along with the decrease from 2019 to 2023 for the R1 and R2 institutions, could indicate that future growth for bepress IRs will not be centered on very high and high research institutions.

The consistent presence of bepress among all R1 and R2 institutions (82 of 279 in 2023, 29.39%) is also worth discussion. The overall market for IRs, and therefore economic sway, is constrained by the number of institutions that exist as potential customers. When dealing with R1 and R2 U.S. institutions, this number is well under 300 total. Conversely, future platform changes by even a small number of institutions would make a considerable difference in the bepress portfolio (or that of other repository platform providers). While outside of the research methods established for this study, additional research could focus on repository platform choice through surveys and/or interviews. It would also be interesting to attempt correlation beyond the Carnegie basic classification with platform choice by looking at other institutional measures such as library budget, enrollment, or National Science Foundation Higher Education Research and Development Survey rankings. Longitudinal studies that include other repository platforms is likewise an area for future research.

From 2017 to 2023, collective holdings in this group of bepress R1 and R2 repositories more than doubled from 1.3 million items to greater than 3 million items. The growth rate of the number of items in R1 and R2 bepress IRs certainly exceeds the growth in the number of repositories in the dataset. While there was a significant standard deviation, individual repository growth, in terms of total items, was seen across the entire set of R1 and R2 bepress repositories, from the largest to the smallest. Due to the variance, it is helpful to view the median values in addition to the averages, especially for individual institutional benchmarking. For example, while on average R1s had roughly 20,000 more repository items than R2s, the largest R2 repository (with almost 100,000 items) was more than double the R1 item average and triple the R1 item median.

The growth and variance in the total number of items give some gauge of what type of elasticity and future capacity shared repository infrastructure may need. Precise planning for future needs remains challenging, with both a pattern of doubling growth every six years and a reality in which the addition or removal of a very large institution would have a significant impact on the total size of a shared repository platform. With regard to IR functionality, this pattern of growth should not solely be taken as a gauge for storage needs. It also indicates a need for parallel development of tools and processes needed to manage, disseminate, and preserve this growing body of work and associated metadata. As referenced by several in the literature review, it is important to remember that metrics such as items and downloads are only part of the picture with regard to sustainability. High engagement with the IR may represent both an

increased risk and reward, depending on the readiness of the university and professional community to support growing demand.

The readership impact, as measured by downloads, is sizable, with a growth from 53 million yearly downloads in 2017 to 88 million yearly downloads in 2023 for the R1 and R2 bepress repositories. There was a decrease in total downloads from the 2021 observation of 107 million yearly downloads. The decline is perhaps attributable to the increased need for online materials during the initial years of the COVID-19 pandemic and subsequent return to in-person instruction and research, ongoing changes by bepress to filter out robots, or simply a decline in use. Items from R2 institutions were downloaded at a similar rate as those of the R1 institutions. This potentially demonstrates the importance of contributions from all types of institutions and that each IR shares a part in contributing unique items of wide interest to global readers.

Throughout all the years of observation, theses and dissertations remained an important part of the R1 and R2 repository landscape, especially as measured by downloads per item. For example, in the 2023 observation for those institutions with such data available, theses and dissertations accounted for 16% of the total number of items and a third of all their downloads. In postulating reasons for such ongoing popularity, it could speak to the value of graduate education and the often new niches of inquiry that these students explore. In addition, in most cases outside of the student posting the work themselves or paying an OA fee to a third party, it is most likely that the only openly available copy of their work is in the IR. While downloads are not the only raison d'etre for repositories, one would be hard pressed to say that one-third of repository development efforts are devoted to this body of work. This is especially striking considering some of the specific needs in this area, including approval workflows, embargoes, copyright, supplemental files and data, early career professional identity, downstream likelihood of name changes, institutional requirements, and advisor recognition.

The longitudinal observation on the strong and consistent prevalence of journals is worth discussing in light of IR directions. Journals were found across the entire set, with only one R1 institution not observed using the bepress journal series format. Usage or absence of this feature does not necessarily provide insight into the actual publishing program or services that might be present at these institutions. For specific analysis of library publishing activities since 2014, the annual *Library Publishing Directory* is a good source (Library Publishing Coalition, n.d.). However, this study does raise interesting questions about how integral journal publishing options are to the IR service model. For example, the bepress platform includes journal publishing as part of its basic subscription, and the strong prevalence of journals may indicate that if publishing capabilities exist on an IR platform, they are likely to be used. It is not clear to what degree IR platform choices are informed by the demand of an institution's

journal publishing program or, conversely, to what degree an institution's journal publishing is informed by available tools on its repository platform. Likewise, during the period of the study, bepress expanded offerings with regard to other areas, such as special collections, native streaming, and automated harvesting of articles. While it is not known whether these new services had an impact on the item growth, it is another example of implications for the ongoing evolution of IRs. The range of these new features get to the heart of the question "What *is* an IR?" and speak to shared infrastructure by asking "What do we need an IR to be?" Institutions will need to continually grapple with the tradeoff between a unified repository platform to hold and disseminate a variety of digital assets versus specialized platforms for assets such as journals, data, special collections, and advanced media, always within the context of resource constraints.

CONCLUSION

This study documents IR item size, item additions, and downloads for R1 and R2 institutions using the bepress platform from 2017 to 2023. The study found roughly 1.8 million items added and 416 million new downloads cumulatively over the six-year period for these institutions. The prevalence of journals was found across more than 90% of these repositories in the study, as was the consistent readership popularity of theses and dissertations. Comparisons between the R1 and R2 bepress institutions revealed that, while the R1 institutions had more total items and items added yearly, the downloads per item figure was approximately the same for R1s and R2s. The study also documents the bepress customer base, noting some growth overall as well as various institutions coming and going. For the R1 and R2 bepress institutions, there was growth overall from 76 in 2017 to 82 in 2023, although the peak was in 2019 at 90 total institutions.

Even with previously stated limitations for generalization, by representing close to 30% of all R1 and R2 institutions, this study contributes toward an understanding of the current IR landscape. There is also circumscribed relevance for forecasting IR infrastructure capacity and direction. Given patterns of content acquisition and readership demand found in this study, it is not unreasonable to anticipate the growing need for IRs that are exceedingly nimble, scalable, and adaptable.

REFERENCES

Arlitsch, K., Wheeler, J., Pham, M. T. N., & Parulian, N. N. (2021). An analysis of use and performance data aggregated from 35 institutional repositories. *Online Information Review*, *45*(2), 316–335. https://doi.org/10.1108/OIR-08-2020-0328

Bailey, C. W., Coombs, K., Emery, J., Mitchell, A., Morris, C., Simons, S., & Wright, R. (2006). SPEC kit 292: Institutional repositories. Association of Research Libraries, Washington, DC. https://publications.arl.org/Institutional-Repositories-SPEC-Kit-292/

Bell, J. (2020, January 10). *F1000 Research joins Taylor & Francis Group*. Taylor & Francis Newsroom. Retrieved January 23, 2024, from https://newsroom.taylorandfrancisgroup.com/f1000-research-joins-taylor-francis/

bepress. (n.d.). *Customers*. Retrieved January 30, 2024, from https://bepress.com/products/digital-commons/why-digital-commons/customers/

Bruns, T., & Inefuku, H. W. (2016). Purposeful metrics: Matching institutional repository metrics to purpose and audience. In *Making institutional repositories work* (pp. 213–234). Purdue University Press. https://doi.org/10.2307/j.ctt1wf4drg.21

COPIM (Community-led Open Publication Infrastructures for Monographs). (2021). *COPIM statement on the corporate acquisition of OA infrastructure*. Copim. https://doi.org/10.21428/785a6451.123ec90e

Dohe, K. (2019). Care, code, and digital libraries: Embracing critical practice in digital library communities. *In The Library With The Lead Pipe*. Retrieved January 29, 2024, from https://www.inthelibrarywiththeleadpipe.org/2019/digital-libraries-critical-practice-in-communities/

Dubinsky, E. (2014). A current snapshot of institutional repositories: Growth rate, disciplinary content and faculty contributions. *Journal of Librarianship and Scholarly Communication*, *2*(3). https://doi.org/10.7710/2162-3309.1167

Fister, B. (2017, August 8). *Just business*. Inside Higher Ed. Retrieved January 30, 2024, from https://www.insidehighered.com/blogs/library-babel-fish/just-business

Ghaphery, J., Childs, A., & Bjork, K. (2024). 2023 *Analysis of 134 bepress Digital Commons research institutions*. https://doi.org/10.17605/OSF.IO/VU698

Goudarzi, S., & Dunks, R. (2023, June 21). Defining open scholarly infrastructure: A review of relevant literature. Zenodo. https://doi.org/10.5281/ZENODO.8064102

Jisc. (n.d. [a]). *IRUS*. Retrieved January 29, 2024, from https://irus.jisc.ac.uk/r5/

Jisc. (n.d. [b]). *OpenDOAR statistics*. Retrieved January 29, 2024, from https://v2.sherpa.ac.uk/view/repository_visualisations/1.html

Knowledge Unlatched. (2021, December 2). *Wiley acquires open access innovator Knowledge Unlatched*. Knowledge Unlatched. Retrieved January 23, 2024, from https://knowledgeunlatched.org/2021/12/wiley-acquires-oa-innovator-ku/

Lewis, D. W. (2017). The 2.5% commitment [working paper]. ScholarWorks Indianapolis. https://doi.org/10.7912/C2JD29

Lynch, C. A., & Lippincott, J. K. (2005). *Institutional repository deployment in the United States as of early 2005*. D-Lib Magazine. *11*(09). https://doi.org/10.1045/september2005-lynch

Library Publishing Coalition. (n.d.) *Library Publishing Directory*. Retrieved February 2, 2024, from https://librarypublishing.org/lp-directory/

Lyrasis. (n.d.). *It takes a village: Open source software models of collaboration & sustainability*. Retrieved January 29, 2024, from https://www.lyrasis.org/programs/Pages/IMLS-OSS.aspx

McDowell, C. S. (2007). *Evaluating institutional repository deployment in American academe since early 2005: Repositories by the numbers, Part 2*. D-Lib Magazine. *13*(9/10). https://www.dlib.org/dlib/september07/mcdowell/09mcdowell.html

Neylon, C. (2018, January 5). *Against the 2.5% commitment*. Science in the Open: The Online Home of Cameron Neylon. Retrieved January 29, 2024, from https://cameronneylon.net/blog/against-the-2-5-commitment/

Petrowski, M. J. (2015, September 15). *Introducing the new ACRL annual survey—Opens September 18*. ACRL Insider. Retrieved January 30, 2024, from https://acrl.ala.org/acrlinsider/introducing-the-new-acrl-annual-survey-opens-september-18/

RAMP (Repositories Analytics and Metrics Portal). (n.d.). Retrieved January 29, 2024, from https://www.lib.montana.edu/ramp/

Skinner, K. (2019). *Mapping the scholarly communication landscape: 2019 census*. Educopia Institute. Retrieved January 30, 2024, from https://educopia.org/wp-content/uploads/2019/06/Census2019_EducopiaPublications.pdf

Tillman, R. K. (2019, July 30). *Repository ouroboros*. Ruth Kitchin Tillman. Retrieved January 29, 2024, from https://ruthtillman.com/post/repository-ouroboros/