

# Who do you love? A report on library investments in scholarly communication infrastructure

November 2020

David W. Lewis and Mike Roy

## Introduction

In the winter of 2020, before the global pandemic, we set out to answer a relatively simple question: How much do libraries spend on the infrastructure that supports publication, access, and preservation of scholarship and data? We were interested in this question as a way to measure in financial terms just how committed libraries are to supporting the work of creating, vetting, distributing, and preserving the intellectual outputs of their institutions.

## Process

In order to arrive at an overall figure, we used a taxonomy developed by CARL that broke expenses down in the following ten categories

1. hosted data platforms
2. hosted publishing
3. hosted repository
4. local data platforms
5. local publishing
6. local repository
7. Preservation
8. consulting services
9. scholarly communication tools
10. staffing

We then developed a survey tool ([link](#)) that allowed libraries to record their investments by each of these categories. Since there was no national list of academic libraries available to us, we distributed the survey through a state, regional, and national academic library groups and consortiums. The web-based survey was open from November 15 through the end of December 2019. We had 91 libraries respond to the survey. The demographics of the respondents are as follows:

- 30 Doctoral Universities with very high research activity
- 24 Baccalaureate Colleges
- 10 Doctoral Universities with high research activity
- 9 Master's Colleges and Universities with larger programs

The remaining 18 libraries are from a wide range of institution types.

## Caveats

Before we report on the findings, we would like to suggest that there are some caveats and issues that make drawing any major conclusions from this survey difficult. We will discuss in the recommendations section of this report ideas we have for addressing the issues that arose in our process.

- Low response rate: With only 91 libraries reporting, we do not think that this is a sufficiently large data set to be able to draw any firm conclusions from.
- Bias towards “fans” of open: We suspect that there was a bias towards “open” and that with more respondents, the results might skew less towards investment in open access and open technologies.
- Some of the data is incomplete: Not all 91 libraries completed the entire survey, and so in some cases, some of the results represent an even smaller sample size. One of the reasons that not all of the data was completely reported is that many of the respondents reported that it was very challenging to locate the data that we were asking for. We’ll comment on that in our analysis.
- Much investment takes place outside of the library: Constraining the investor to the library only skewed the results as many schools fund this via multiple areas of the institution (e.g. IT, Provost).

## Findings

What follows are some high level charts that provide a view into the data.

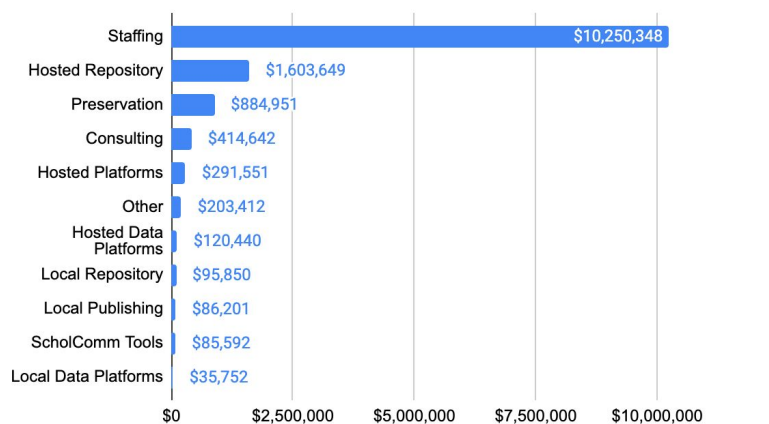
### Top Level Summary

The 91 respondents collectively invested **\$14.1 million** in this area, broken down as follows:

- \$150,000 average per library
- \$8 per student enrolled
- 1.5% of overall library budget (including salaries)
- 2.8% of library budget (excluding salaries)

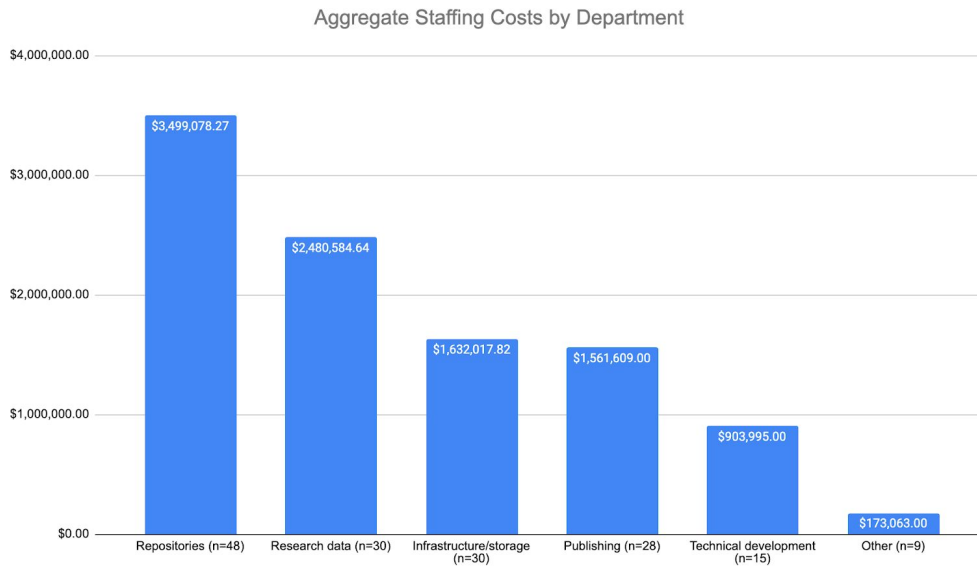
### Investment by Category

Aggregate Investment by SCR Type



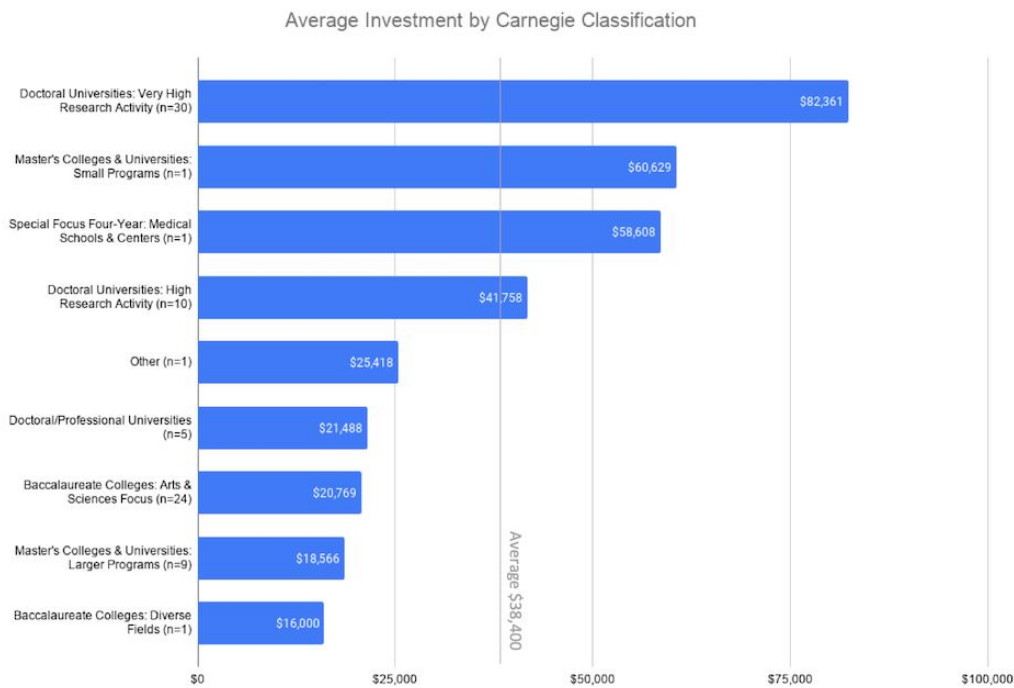
As is made clear by this graph by far the largest part of the investment, over 75% of total investment, went into staffing.

# Staffing Costs by Area



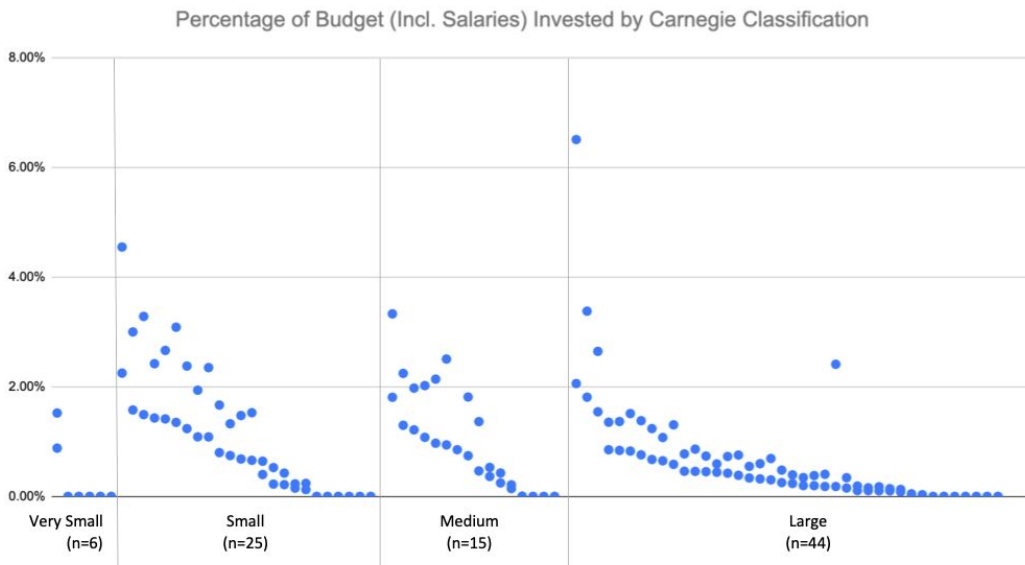
In analyzing staff costs, we found that the majority of staffing is going to support repositories, followed by research data.

# Investment by Carnegie Class



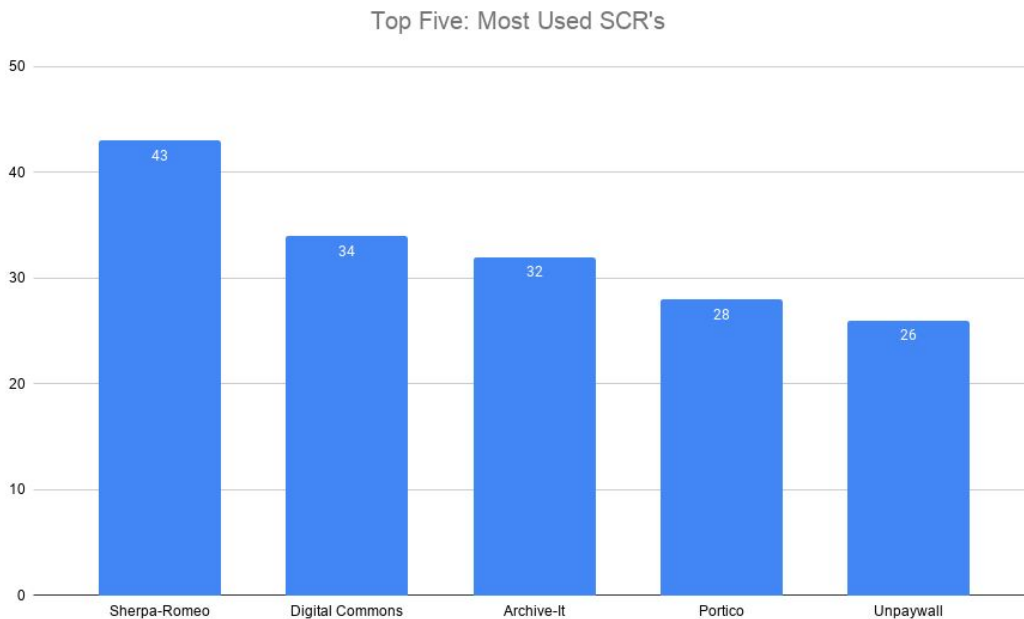
Unsurprisingly, we found that larger schools with larger budgets invest more than smaller schools with smaller budgets.

## Investment as Percentage of Budget



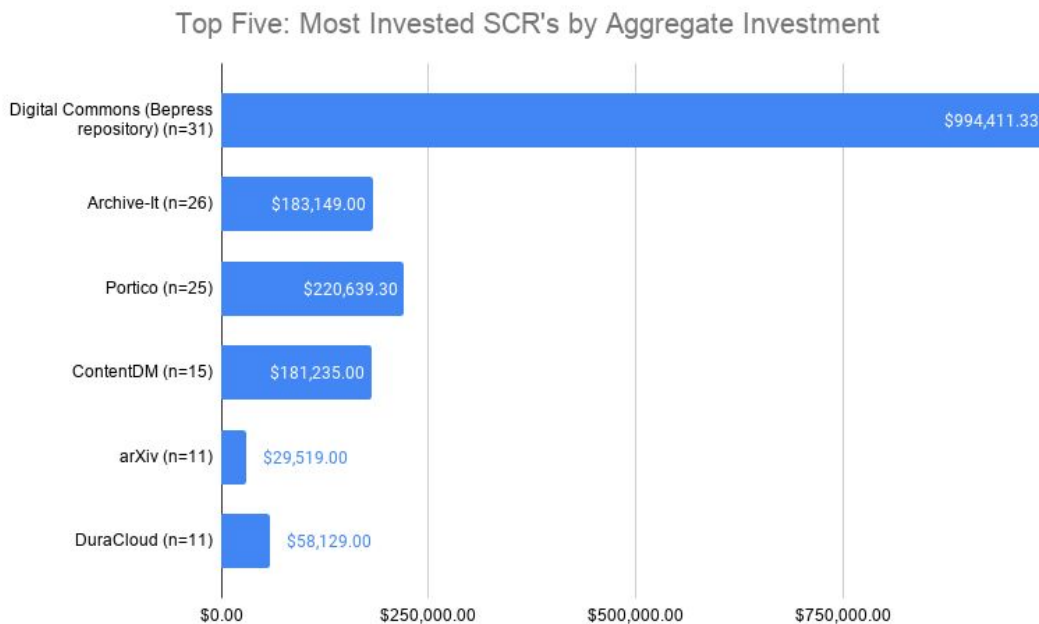
However, we found that regardless of school size, the majority of respondents were spending less than 2% of their total budget on this infrastructure.

## Most Used SCR's



For the 91 libraries, there were 490 “uses” of SCR's reported, meaning that on average, each library on average used just over 5 SCR's. Out of a total of 57 unique SCR's reported on, these are the top five in terms of number of usages reported.

## Most Invested SCRs



The 91 respondents reported a total of 212 investments in SCRs, meaning that each library invested on average in just over 2.3 SCRs. These are the five SCRs that had the largest number of libraries investing in them. Total investment by the 91 libraries was \$2.6 million. These are the SCRs that received the largest amount of investment in terms of dollars, and also in terms of the number of libraries making investments of any amount.

## Summary Findings

- The average investment by the participating libraries is \$150,000, which is less than 2% of total library spending, and represents about \$8 per enrolled student.
- Larger schools invest more per school, but about the same as smaller schools as a percent of total budget.
- Nearly 75% of all investment goes into staffing.
- While certain SCRs attracted a great deal of users and investment, the majority of investments went to a broad array of SCRs.
- Many of the most used SCRs were not receiving commensurate investments given the level of use. That is, many libraries are free riding on the investments of others.

## Conclusions and Recommendations

In looking at the data reported, and reading the comments from our respondents, it is clear that our ability to easily locate and track over time investments in this infrastructure is not yet fully developed. Libraries have spent decades developing sophisticated methods for tracking a wide range of statistics about our services and our expenditures, but it is very early days in terms of our ability to understand this dimension of our work. Most

academic libraries have only a vague understanding of their investments in the infrastructure that underpins their digital collections and services. The instrument we developed to capture this information is a good rough draft, but will require refinement to better map onto how libraries track their spending. There is also work to be done to build into our individual accounting practices ways to flag these expenses to make the process of tracking expenses far easier. On a macro level, national level effort to establish standard definitions and standard ways of measuring investments is required if this effort is to be successful. In addition, on-going maintenance of this data will require that some existing library organization see the value in maintaining this data. Absent this, we will have difficulty understanding the changing nature of academic library practice.

The data represents a particular moment in time, namely the winter of 2019. The global pandemic makes this data of particular interest should a similar study be conducted in the next few years, as it would allow us to see how libraries as they recover from this pandemic have prioritized investments in these infrastructures.

## **Acknowledgments**

We would like to thank the following for their good advice and contribution to this effort: The Canadian Association of Research Libraries, and in particular, Kathleen Shearer, for sharing their survey instrument. True Bearing, and Nathan Brown and Brianna Morrow in particular, for their expert advice on survey design. Our Advisory Committee for their help in encouraging participation. And of course to the 91 libraries that contributed their data to this effort.