



TEXAS TECH UNIVERSITY  
Libraries™

**RESULTS FROM A SECOND LONGITUDINAL SURVEY OF ACADEMIC RESEARCH  
COMPUTING AND DATA CENTER USAGE: EXPENDITURES, UTILIZATION PATTERNS,  
AND APPROACHES TO RETURN ON INVESTMENT**

The Texas Tech community has made this publication openly available. [Please share](#) how this access benefits you. Your story matters to us.

Citation	Sharon Broude Geva, Alan Chalker, Curtis W. Hillegas, Donald Petravick, Alan Sill, and Craig A. Stewart. 2021. Results from a second longitudinal survey of academic research computing and data center usage: expenditures, utilization patterns, and approaches to return on investment. In Practice and Experience in Advanced Research Computing (PEARC '21), July 18–22, 2021, Boston, MA, USA. ACM, New York, NY, USA, 4 pages. <a href="https://doi.org/10.1145/3437359.3465589">https://doi.org/10.1145/3437359.3465589</a>
Citable Link	<a href="https://hdl.handle.net/2346/90392">https://hdl.handle.net/2346/90392</a>
Terms of Use	<a href="#">CC BY 4.0</a>

# Results from a second longitudinal survey of academic research computing and data center usage: expenditures, utilization patterns, and approaches to return on investment

Sharon Broude Geva\*  
sgeva@umich.edu  
University of Michigan  
Ann Arbor, Michigan, USA

Alan Chalker\*  
alanc@osc.edu  
Ohio Supercomputer Center  
Columbus, Ohio, USA

Curtis W. Hillegas\*  
curt@princeton.edu  
Princeton University  
Princeton, New Jersey, USA

Donald Petravick\*  
petravick@illinois.edu  
University of Illinois  
Illinois, USA

Alan Sill\*  
Alan.Sill@ttu.edu  
Texas Tech University  
Lubbock, Texas, USA

Craig A. Stewart\*  
stewart@iu.edu  
Indiana University; University of  
Illinois  
Urbana-Champaign, Illinois, USA

## ABSTRACT

Availability of cloud-based resource delivery modes is transforming many areas of computing. Academic research computing and data (RCD) support largely remains based on on-premises delivery and has adopted commercial clouds more slowly than the private sector for a variety of stated reasons including factors related to cost efficiency, return on investment, institutional requirements, high costs for bulk commercial cloud computing usage, and funding patterns. Other factors involved in selection of computing resource delivery modes include capabilities and applications that are available only in or best adapted to specific computing environments. It is important for the higher education and research communities to be able to learn from each other as institutions and individuals to make optimum use of appropriate modes of delivery for RCD resources. This paper reports an overview of results from the second annual community-wide survey conducted by the Coalition for Advanced Scientific Computation on patterns of funding, usage, and return on investment for academic research computing and data resources. The results show that on-premises delivery continues to remain the preferred mode for RCD resources for most responding institutions as found in the first survey, but that commercial cloud usage is beginning to be reported for production use by a small number of respondents to the survey. Reasons for these preferences are further explored in the survey and initial high-level results are reported here.

## CCS CONCEPTS

• **Social and professional topics** → **Computing / technology policy.**

\*All authors contributed equally to this research.



This work is licensed under a Creative Commons Attribution International 4.0 License.

PEARC '21, July 18–22, 2021, Boston, MA, USA  
© 2021 Copyright held by the owner/author(s).  
ACM ISBN 978-1-4503-8292-2/21/07.  
<https://doi.org/10.1145/3437359.3465589>

## KEYWORDS

cloud computing, HPC, research computing, data-centric computing, return on investment

### ACM Reference Format:

Sharon Broude Geva, Alan Chalker, Curtis W. Hillegas, Donald Petravick, Alan Sill, and Craig A. Stewart. 2021. Results from a second longitudinal survey of academic research computing and data center usage: expenditures, utilization patterns, and approaches to return on investment. In *Practice and Experience in Advanced Research Computing (PEARC '21)*, July 18–22, 2021, Boston, MA, USA. ACM, New York, NY, USA, 4 pages. <https://doi.org/10.1145/3437359.3465589>

## 1 INTRODUCTION

The large majority of disciplines and sub-disciplines within the general area of scientific and engineering research are dependent upon research computing and data analysis (RCD) facilities. Trade-offs in selecting the proper mode of delivery include considerations related to cost efficiency, return on investment (ROI), convenience for researchers, staff, faculty, and students making use of RCD resources, time to delivery for solutions, availability of software, data, and usage expertise, and many other factors.

The Coalition for Academic Scientific Computation (casc.org) is a not-for-profit 501(c)(3) organization founded in 1989 with over 90 member institutions that include many leading organizations in the US active in the areas of research computing, cyberinfrastructure, and data analysis and several similarly-situated non-US international partners. Sharing best practices in RCD support is part of CASC's mission. CASC formed a working group in 2018 to study and disseminate information about academic research computing and data facilities with a charter to collect and share information about ROI for academic usage of data centers and clouds. This group conducted its first studies of commercial cloud and on-premises facilities usage in the form of a short survey in late calendar year 2019 to determine topics of interest followed by a more detailed one in early 2020, and reported the results of these surveys at PEARC'20[2].

Continued interest in the area of ROI comparisons for on-premises vs. cloud computing was evidenced by the response to these surveys and by strong attendance at an open community workshop on ROI

**Table 1: Number of Respondents by Compute Capacity for On-premises and Commercial Clouds**

Total number of CPU cores available across all on-premises computational resources	0-100	100-1K	1,000-10K	10,000-100K	>100K
	2	8	18	34	8
Total number of CPU core hours utilized by researchers on commercial cloud computational resources	0-100K	100K-1M	1M-10M	10M-100M	>100M
	43	10	6	4	1

for research computing and data infrastructure held in fall 2020, which drew a total of 107 participants. Of the workshop participants, 26 were from organizations that were not CASC members, including two from Europe and four from industry/non-profit organizations. Based on considerable support to continue this activity from participants at this workshop, CASC engaged in its second annual community-wide survey of resource funding and usage for academic research computing and data support. Among the points of consensus at the workshop was that it is now more than ever important to understand how academic institutions plan, budget for, and use research computing and data facilities because of changes rapidly taking place in academia such as those due to COVID-19 as well as demographic changes that are creating financial challenges for institutions of higher education due to dramatically changing work and learning patterns.

The purpose of this brief paper is to present the most critical results from the most recent 2021 survey in comparison to 2020 survey results. A longer report with more details of methods and analysis will be published in the near future.

## 2 METHODS

The methods for this survey were generally similar to last year’s survey as described in [2]. In response to feedback from participants in that survey, we created a shorter survey instrument this year. It consisted of a brief introduction, along with a list of common terms and their definitions (e.g. Research Computing and Data), followed by six sections, each with five to seven questions. The sections thematically grouped questions into the following topics: Demographics, Resources, Investments, Return on Investments, Policies and Procedures, and Miscellaneous.

The survey was administered via the Qualtrics online survey system. Respondents were assured that individual responses would remain anonymous, but identifying information was requested in order to assure unique responses by institution. We issued a broad appeal for participation to several academic computing groups and organizations within the US and internationally and specifically asked for assistance in obtaining broad dissemination to improve the breadth and diversity of representation of participating organizations.

While CASC member institutions were asked to indicate their membership, this was not required to participate in the survey.

We also simplified the format for many questions to make the survey easier to fill out and more compact. When asking about usage of various forms of computing and data resources, for example, we typically asked respondents to characterize their use of a particular type of resource as "predominant," "significant," or "Minimal/None." The majority of the questions only required the respondent to select from a short list of possible answers. A few questions asked for

short text answers, and only the first few demographic questions were required to be answered. This simplified both responding to the survey and analysis of the results.

The survey collected data between February 12 and March 16, 2021. We received a total of 71 responses, of which 69 were substantially complete and 52 were from CASC members. The full survey form and a more complete explanation of survey results is available online [1]. The most critical results from this survey are summarized in the tables and discussion below.

## 3 RESULTS AND DISCUSSION

There are several changes between the pattern of results from early 2020 and those from the current 2021 survey. Before discussing these changes, we note that while the overall number of responses is greater this time (69 compared with 53 previous completed responses), only 39% of the current respondents indicated they had participated in the previous year’s surveys. This could be partly due to the greater emphasis placed this year on obtaining responses that were authoritative: 61 out of 69 completed responses indicated that they were responding authoritatively on behalf of their institution, whereas this factor was not measured in previous surveys.

Table 1 shows the number of responses in terms of CPU capacity for both on-premises computational resources and those purchased from commercial cloud sources. Note that the bins were chosen for numerical convenience and are not directly comparable: 100 cores of on-premises computation could deliver as many as 872K CPU core-hours per year if fully utilized, or 785K core-hours at 90% utilization, but as cloud use is just beginning to be seen at scale for academic RCD usage, the lower-valued bin limits allow us to see this onset with more sensitivity to the low end of the range.

The table shows that the vast majority of institutions are only making limited use of commercial cloud computational resources, at the scale of a single modern 128-core server or so, but a small number of institutions report up to 100x such usage or more. In contrast, the majority of respondents (42/69) reported on-premises resources in excess of 10,000 cpu cores, capable of delivering more than 78.5M CPU core-hours at 90% utilization.

In 2020, no respondent indicated that commercial cloud services were a major resource for their institution, and no respondent indicated that their institution was pursuing use of cloud resources in order to save money. The 2021 survey asked questions designed to expand on the details of service usage and delivery and relative preference for computational and storage resources.

One question asked in 2021 probed the relative predominance of the use of various forms of resource delivery. While the question was designed to determine what method of delivery was used more than others at the institution, it seems that most respondents interpreted it to be asking whether that method was used by a

large portion of their research community. In the current 2021 survey responses, as shown in Figure 1, two respondents indicated cloud storage was a predominant source of data storage for their institution.

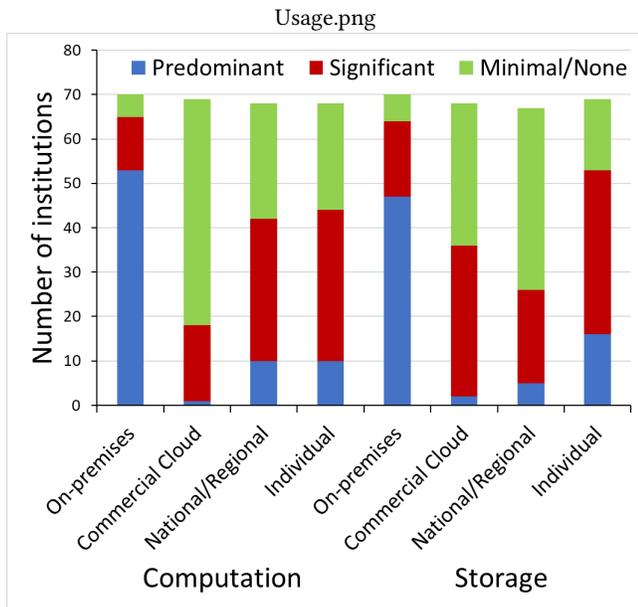


Figure 1: Relative level of usage by researchers organized by mode of delivery.

While more institutions indicated that on-premises resources were either predominant or significant in terms of utilization by researchers (63 out of 69 responses for computation and 62 out of 69 for storage), nearly half of respondents also indicated significant or predominant utilization by researchers for commercial cloud resources (16 out of 69 for cloud computation, and 36 out of 69 for cloud storage).

Another question asked respondents to rate their assessment of the relative return on investment for various forms of delivery. This question was more direct in terms of asking for the value of the delivery method to that institution in their research setting. Figure 2 shows the percentage of respondents in terms of their assessment to this ROI question, binned by the respondents' response to a different question about their institutions total supported research income in order to group similarly sized institutions. A very large majority of respondents indicated that on-premises resources represented better ROI for meeting the majority or all of institution's research computing needs, but this year, two responded that commercial cloud is better for more situations than on-premises resources at their institutions. While neither of these institutions had participated in the 2020 surveys, this is a significant change from the previous year's results.

A significant fraction of the responses in each supported research category indicated that on-premises resources were better in terms of ROI than commercial cloud for more situations, but did not choose the "nearly all" response, indicating that institutions are finding value in use of commercial cloud resources in at least some

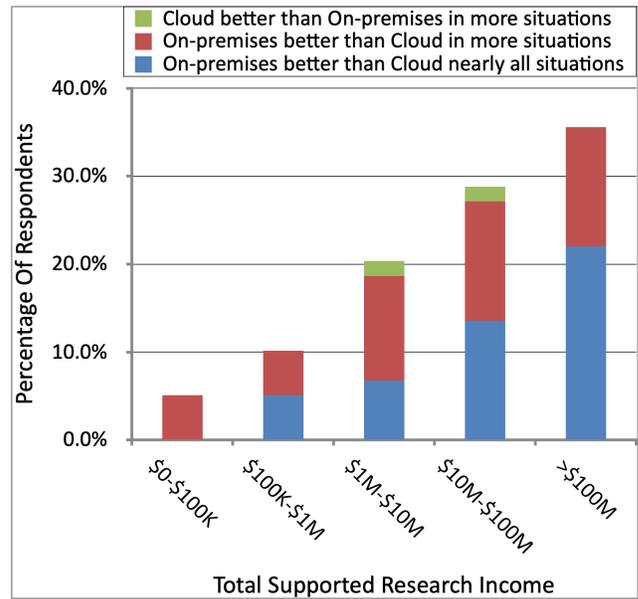


Figure 2: Percentage of total respondents to the 2021 survey to the question regarding assessment of ROI depending on the mode of delivery and level of supported research income.

situations. There were still no responses this year that indicated commercial cloud to be better in nearly all situations than on-premises resources for RCD needs.

The above results are influenced by patterns observable in funding models for RCD resources as shown in Table 2. Many institutions make use of some sort of one-time or intermittent pattern to finance RCD resources, which is inconsistent with the continuous funding model needed to make use of commercial cloud resources. As noted in last year's report and in other reports, one of the obstacles to adoption of commercial cloud resources in academia is the shift in fiscal model from periodic one-time investments to ongoing operational costs that is required for such a change.

The above findings show clear conclusions that cost savings are not the primary driver of academic RCD interest in commercial cloud services. Most respondents, although not all, in this year's survey continue to prefer on-premises delivery when assessing ROI patterns. Similar concerns about long term costs of commercial cloud resources have been expressed in cost analyses for the private sector [5]. As was also seen in the 2020 CASC surveys, however, institutions nevertheless continue to explore and probe the use of commercial cloud-based resources for computation and storage.

The most common reason for exploration of commercial cloud resources was given as "experimentation with utility for research", followed fairly distantly by a cluster of three answers: "need for cloud-native capabilities" then "autonomy for researchers to make their own choices" and "need for 'always on' capability." Significantly different from last year, 10 institutions report making use of commercial cloud resources in order to save money at least in some circumstances. The data continue to show a trend towards

**Table 2: Funding models for Computing Resources**

What funding models are used for capital resources/operational resources (staff and support) for RCD? (Select all that apply)	Capital	Operational
The institution makes one-time or intermittent investments in on-premises resources	36	n/a
The institution provides a regular budget for on-premises resources	35	37
Colleges or departments fund on-premises resources that are offered to researchers within their own organizational unit	39	29
Researchers or research groups write grant requests to fund resources	55	25

**Table 3: Organizational Units that compile RCD Data**

Levels at which data on RCD-related topics are compiled	Institution	College/Dept	Researchers
RCD Resource Usage	41	42	54
Academic Outcomes	33	29	31
Return on Investment (ROI)	32	19	16

higher rates of pilot cloud exploration projects among institutions with larger amounts of supported research funding.

The survey probed the level at which respondents document the impact of their RCD support in both non-financial[3] and financial[4] terms independent of the mode of delivery. As shown in Table 3, the fractions of institutions calculating and reporting RCD resource usage (41 out of 68), academic output impact (33 out of 69), and ROI on RCD resources to upper administration (32 out of 69), averaging around 50%, are higher than in last year's survey in which they averaged about 30%, but remain far from ubiquitous. In response to the question on what improvements would help expand access to academic RCD resources, "increased federal funding for such resources" was the most common answer. A majority of respondents also indicated "resources to help with making the case for better internal support for RCD within our institution" were needed to expand access to academic RCD resources.

#### 4 CONCLUSION AND FUTURE PLANS

This second annual CASC survey showed slight but significant changes from last year's survey in terms of exploration and adoption of commercial cloud resources. On the whole, results remain largely consistent with last year's results. On-premises delivery of RCD resources remains by far the preferred method in terms of ROI at the vast majority of responding institutions. This year, however, for the first time, a few institutions indicated that commercial cloud resources had become a useful form of delivery for at least some situations their research community and a small number of respondents indicated cloud resources to be better in terms of ROI than fully on-premises delivery in at least some circumstances. While no respondent last year indicated that cost savings were achieved by use of commercial cloud resources, this year 10 responses indicated cost savings as a driver for at least some commercial cloud adoption. Experimentation with commercial cloud capabilities and use of cloud-native capabilities were the two most common drivers

for use of commercial cloud resources for those institutions making use of this form of delivery.

CASC plans to continue this survey on an annual basis to continue to track trends and ROI assessments in the academic RCD community. Additional information on methodology and a more extensive presentation of results will be made available in a future publication.

#### ACKNOWLEDGMENTS

The authors thank the CASC Executive Officers and every person who submitted responses to the survey.

#### REFERENCES

- [1] Sharon Broude Geva, Alan Chalker, Curtis W. Hillegas, Alan Sill, Craig A. Stewart, and Donald Petravick. 2021. Survey form and methods for second CASC survey of academic research computing and data center usage. <https://doi.org/10.7302/1271>
- [2] Alan Chalker, Curtis W. Hillegas, Alan Sill, Sharon Broude Geva, and Craig A. Stewart. 2020. Cloud and On-Premises Data Center Usage, Expenditures, and Approaches to Return on Investment: A Survey of Academic Research Computing Organizations. In *Practice and Experience in Advanced Research Computing* (Portland, OR, USA) (PEARC '20). Association for Computing Machinery, New York, NY, USA, 26–33. <https://doi.org/10.1145/3311790.3396642>
- [3] Craig A Stewart, Amy Apon, David Y Hancock, Thomas Furlani, Alan Sill, Julie Wernert, David Lifka, Nicholas Berente, Thomas Cheatham, and Shawn D Slavin. 2019. Assessment of Non-financial Returns on Cyberinfrastructure: A Survey of Current Methods. In *Proceedings of the Humans in the Loop: Enabling and Facilitating Research on Cloud Computing* (Chicago, IL, USA) (HARC '19). ACM, New York, NY, USA, 2:1–2:10. <https://doi.org/10.1145/3355738.3355749>
- [4] Craig A. Stewart, David Y. Hancock, Julie Wernert, Thomas Furlani, David Lifka, Alan Sill, Nicholas Berente, Donald F. McMullen, Thomas Cheatham, Amy Apon, Ron Payne, and Shawn D. Slavin. 2019. Assessment of Financial Returns on Investments in Cyberinfrastructure Facilities: A Survey of Current Methods. In *Proceedings of the Practice and Experience in Advanced Research Computing on Rise of the Machines (Learning)* (Chicago, IL, USA) (PEARC '19). ACM, New York, NY, USA, Article 33, 8 pages. <https://doi.org/10.1145/3332186.3332228>
- [5] S. Wang and M. Casado. 2021. The Cost of Cloud, a Trillion Dollar Paradox. Andreessen Horowitz (a16z). <https://a16z.com/2021/05/27/cost-of-cloud-paradox-market-cap-cloud-lifecycle-scale-growth-repatriation-optimization/>