




# Evaluating Return on Investment for Cyberinfrastructure Using the International Integrated Reporting <IR> Framework

Winona G. Snapp-Childs<sup>1,2</sup> · David L. Hart<sup>3</sup> · Claudia M. Costa<sup>2</sup> · Julie A. Wernert<sup>1</sup> · Harmony E. Jankowski<sup>1</sup> · John W. Towns<sup>4</sup> · Craig A. Stewart<sup>5</sup> 

Received: 1 October 2023 / Accepted: 10 April 2024  
© The Author(s) 2024

## Abstract

This paper investigates the return on investment (ROI) in cyberinfrastructure (CI) facilities and services by comparing the value of end products created to the cost of operations. We assessed the cost of a US CI facility called XSEDE and the value of the end products created using this facility, categorizing end products according to the International Integrated Reporting Framework. The US federal government invested approximately \$0.3B in operating the XSEDE ecosystem from 2016–2022. The estimated value of end products facilitated by XSEDE ranges from around \$4.7B to \$22.7B or more. Credit for the majority of these end products is shared among various contributors, including the XSEDE ecosystem. Granting the XSEDE ecosystem a seemingly reasonable percentage of credit for its contributions to end product creation suggests that the return on federal investment in the XSEDE ecosystem, in terms of value of end products created, was greater than one and possibly far greater than one. The Framework proved useful for addressing this question. Earlier work showed that the value of services provided by XSEDE was significantly greater than the cost of those services to the US federal government—a positive return on investment for delivery of services. Analyzing the financial efficiency of operations and the financial value of end products are two means for assessing the success of CI facilities in financial terms. Financial analyses should be used as one of many approaches for evaluating the success of CI facilities.

**Keywords** ROI · Return on investment · Cyberinfrastructure · XSEDE · Science gateways · IR framework

**JEL Classification** I23 · M15 · O31 · O32 · O38

**Mathematics Subject Classification** 68U99 · 91B44 · 91B18

This article is part of the topical collection “Metrics for Measuring Success of CyberInfrastructure (CI) Projects” guest edited by Ritu Arora and Amit Majumdar.

✉ Winona G. Snapp-Childs  
wsnappch@iu.edu

David L. Hart  
dhart@ucar.edu

Claudia M. Costa  
clcosta@iu.edu

Julie A. Wernert  
jwernert@iu.edu

Harmony E. Jankowski  
hjankows@iu.edu

John W. Towns  
jtowns@illinois.edu

Craig A. Stewart  
stewart@iu.edu

<sup>1</sup> Pervasive Technology Institute, Indiana University, 2709 E. Tenth St., Bloomington, IN 47408, USA

<sup>2</sup> Research Technologies, University Information Technology Services, Indiana University, 2709 E. Tenth St., Bloomington, IN 47408, USA

<sup>3</sup> Computational and Information Systems Lab, National Center for Atmospheric Research, PO Box 3000, Boulder, CO 803807, USA

<sup>4</sup> National Center for Supercomputing Applications, University of Illinois, 1205 W. Clark St., Urbana, IL 61801, USA

<sup>5</sup> Department of Computer Science, Indiana University, 700 N. Woodlawn Ave., Bloomington, IN 47408, USA

## Introduction

The scientific research community is faced with a myriad of potentially valuable and interesting research questions, with time and funding available to examine only a limited subset of them. Cyberinfrastructure (CI) facilities supporting research generally face more demand for their resources and services than they are able to fulfill, given the time and funding available. Presently, government research funding and university budgets are under significant strain, at least within the US. This leads to questions at all levels, from the federal government to individual labs, about how many and which research projects to fund, and how much funding to expend on providing research-supporting CI facilities and services. For many years now, researchers and CI facility operators have justified their research and research facilities based on intellectual outcomes, the practical effects of those outcomes, or the inherent value of understanding our world. Improved knowledge and understanding are indeed the primary motivations for students and those pursuing careers in research and research support, but sometimes the value of research is most effectively demonstrated in financial terms. For audiences such as chief financial officers (CFOs), elected government representatives, or taxpayers, financial terms are an essential part of any explanation of the value of any scientific research endeavor or service. This paper uses a recently developed accounting methodology, called the <IR> Framework, to analyze the financial value of the end products created via the use of research CI facilities. This methodology allows us to evaluate return on investment (ROI) in a CI facility by comparing the monetary value of the end products created using the facility with the facility's cost, something that has never been done before in the US, as far as we can ascertain.

It has long been possible to relate the value of investment in information technology (IT) to the value of very broad outcomes and end products. A recent example is [1], which shows that what we know as the modern telecommunications industry is a return on government investment in computer science research. This approach, while compelling, cannot help us evaluate the success of a specific research or research infrastructure investment. The challenges in attaching the financial values of end products to investment in a specific research project and a specific research CI facility include:

- *The values of end products accrue over long periods of time and are time-dependent.* For example, chemical simulations using CI resources may produce results that take years, even decades, to come to market as a drug. Once available for use, however, the drug's value

increases each year as it is used to treat more and more people.

- *A CI facility used to support research is only one of many factors that goes into creating a useful end product.* For most end products developed using a CI facility, many other facilities and large numbers of people contribute to the final product. What portion of its value should be credited to the CI facility?
- *It is very hard to value outcomes.* How does one assess the long-term financial value of something as routine as a scientific publication or as extraordinary as successfully mitigating the negative effects of global climate change?

These challenges are all significant. The best available methods for assessing the financial value of CI facilities are unquestionably imperfect. However, in the current fiscal environment, we deem it better to do an imperfect analysis with clearly understood shortcomings than to do nothing. Examples of questions the authors of this paper have heard include (paraphrased): “Our university is in dire financial straits; please explain, in specific financial terms, why your CI facility should not be closed,” and “Funding agency XYZ has a very limited budget; please explain to these congressional staff members why investing in your particular CI facility is of financial benefit to the US economy.” In the case of the eXtreme Science and Engineering Discovery Environment (XSEDE) project, discussed here, the ability to offer a quick, reasonable answer about XSEDE's financial effectiveness was a critical factor in the facility being renewed for the maximum time permitted by the policies of the National Science Foundation (NSF)—the branch of the US federal government that funded the project. Conversely, the authors know of three specific cases where less clear answers about the financial effectiveness of investing in CI facilities were a factor in a facility's downsizing or elimination. One publicly documented example is the closure of the Arctic Region Supercomputer Center (ARSC) [2], part of the University of Alaska Fairbanks (UAF) and a long-time excellent and internationally recognized supercomputing center. Formerly funded largely through the US Department of Defense (DOD), ARSC was closed in 2015 when full financial responsibility for its operations fell to UAF. As with the closure of any university subunit, many factors were at play. In this case, important factors included UAF budget constraints and what its leadership saw as inadequate enunciation of the value proposition of ARSC for the university. (Co-author Stewart was on an external review committee that was unable to explain ARSC's value to UAF to the satisfaction of its institutional leaders.) The financial stresses in higher education are most severe in smaller institutions [3] such as UAF, but they are felt across the higher education landscape. In

the months preceding the writing of this paper, two large research universities announced budget shortfalls in the several tens of millions of dollars, while a third university announced a shortfall exceeding \$10M [4]. None of these has yet involved any announcements of cuts to research CI facilities or support, but one of the situations is so severe that most aspects of university operations are likely to be affected. Other large research universities have downsized their CI facilities in recent years. Thus, there is good evidence that at some universities, investment in CI facilities has been or will be decreased as a consequence of financial challenges (see [5]). A smaller number of universities, meanwhile, have seemingly justified increasing investment in CI facilities. All of this suggests that it is timely to pursue the financial aspects of investment in CI facilities.

In this paper, we analyze the long-term financial value created as a result of US federal government investments in the suite of CI projects collectively called the XSEDE ecosystem [6]. This ecosystem was supported by two different series of grant awards from the NSF. The XSEDE project was funded by one of these series of awards—a series of two awards that funded XSEDE to provide services to the US research and education community starting in 2011 and ending in 2022. The XSEDE award supported centralized allocation, operation, security, training, and support functions for a set of large-scale computation, data storage, and visualization systems. The other series of grant awards by the NSF supporting the XSEDE ecosystem—separate and distinct from the XSEDE project funding—funded construction and operation of several large-scale CI systems, with roughly half a dozen such systems in operation at any given time. Recipients of these grants were required to build CI facilities that made use of XSEDE’s operational and support services and complied with XSEDE technical standards. The NSF awards that funded these large-scale CI systems operated on staggered timescales, allowing the NSF to continually refresh the pool of CI systems it made available to the US research community. The High Performance Computing (HPC) systems integrated with XSEDE included widely used supercomputers and cloud systems such as Stampede, Bridges, and Jetstream. Broadly, the XSEDE ecosystem refers to the NSF-funded project called XSEDE, the additional NSF-funded CI systems, and other smaller systems and support activities not funded by the federal government. To give a sense of the scale of the XSEDE project, at any given time it employed about 180 individuals, representing some 90 full-time equivalent (FTE) staff [7]. In its eleven years of operation, the XSEDE ecosystem supported a total of 32,000 distinct users, located in every state in the US. The XSEDE ecosystem included more than 15 advanced CI systems, with an aggregate processing capability of more than 25 petaFLOPS (1 petaFLOPS = one quadrillion floating-point operations per second). See [7, 8] for more details.

This paper focuses on the time period from September 1, 2016 to August 31, 2022. This was a time of mature operational activities, during which the project collected reliable counts of many types of end products. Given the NSF’s investment of hundreds of millions of dollars in the XSEDE ecosystem for over a decade, it certainly seems worthwhile and responsible to attempt to assess the value of the results of the investment.

We explore the fundamental question regarding the worth of research facilities: Does the expenditure of taxpayer money on the XSEDE ecosystem by the US federal government lead to societal benefits that exceed its costs?

## Prior Related Work

### Assessing Return on Investment in Cyberinfrastructure

Stewart et al. [9] reviewed the literature regarding ROI for cyberinfrastructure. The first peer-reviewed paper to quantitatively analyze the value of financial investment in CI was Apon et al. [10]. They and Smith and Harrell [11] demonstrated that increased investment in CI leads to an increase in desired outcomes including publications, doctoral degrees conferred, and university rankings. Apon et al. and Smith [12] also showed that increased investment in CI leads to increased grant funding.

A textbook definition of ROI is “a ratio relating income generated by an organization to the resources (or asset base) used to produce that income” [13], or:

$$\text{ROI} = \frac{\text{Income}}{\text{Assets invested}}$$

In public sector research, however, there is generally no income per se; nothing is being sold or leased. The term  $\text{ROI}_{\text{proxy}}$  was coined to create a measure conceptually similar to ROI, where value created in financial terms is compared to the financial cost [14]:

$$\text{ROI}_{\text{proxy}} = \frac{\text{Market value of services delivered and products created}}{\text{Cost to deliver services and products}}$$

In this paper, as in earlier work, we refer to the values we estimate as  $\text{ROI}_{\text{proxy}}$  to emphasize that we are using “value created divided by cost” as a proxy for ROI in a situation where no income is actually realized by the entity making the investment. This sort of generalization of the definition of ROI is not new; many prior publications have done this, but without noting that the initial definition of ROI is being adapted rather than followed exactly. Another definition of ROI is profit divided by investment [15]. The biggest difference between these otherwise very similar definitions is that

the break-even point in the definition we use is 1.0, while the break-even point in the definition based on profit is 0.0. Conceptually, XSEDE creates value rather than receiving funds or making a profit. “Value created” seems more akin to “funds received” than to “profit,” and that is the basis for our choice between these very similar definitions.

Our prior work used the logic model of organizational processes shown in Fig. 1 (modified from [16]) as its conceptual framework. A similar approach is taken in assessing the “cost effectiveness ratio” of a diverse set of large research infrastructure facilities in [17].

In earlier works we concentrated on  $ROI_{\text{proxy}}$  for investment in the “Activities” block of this model by comparing the market value of services delivered to the actual cost of delivering those services. These analyses showed that XSEDE delivers services for 1/3 to 1/2 of the cost of purchasing the services from commercial providers: thus, an  $ROI_{\text{proxy}}$  for XSEDE operational activities was between 2 and 3. Similar analyses, comparing purchasing and operating on-premise HPC systems with the cost of commercial cloud services, show that on-premise HPC and cloud facilities are far less expensive (for example, [18]). Prior results *consistently* show that public sector CI services and facilities cost a fraction of what it would cost to purchase those services on the market.

### The International Integrated Reporting <IR> Framework

In this paper we compare the value of end products—outputs and outcomes as shown in Fig. 1—with the cost of operating the XSEDE ecosystem. To categorize and tally the value of end products enabled by the use of XSEDE, we have adopted a framework called the *International Integrated Reporting <IR> Framework*, or <IR> Framework [19]. This framework was created and is maintained by the International Integrated Financial Reporting Standards (IFRS) Foundation, whose website states: the “purpose of an integrated report is to explain to providers of financial capital how an organization creates, preserves, or erodes value over time.” An integrated report benefits all stakeholders, including “legislators, regulators and policy-makers.” The <IR> Framework recognizes six categories of value

creation: financial, manufactured, intellectual, human, social and relational, and natural, as shown in Fig. 2. It seems to account for every sort of tangible value one can imagine as an end product created or enabled by any facility. Using this international standard for analyzing value has the benefit of enabling comparison with the analyses of other facilities that have been carried out using this same framework.

### Methods

As stated previously, the time period analyzed in this study is the final six years of XSEDE operations: September 1, 2016 to August 31, 2022, in other words FY2017 through FY2022 of the second of the two NSF grant awards supporting XSEDE and referred to, within the NSF, as “XSEDE2” [7]. We calculate the cost and value of benefits for the XSEDE ecosystem for this entire time period. All financial values given in this paper are in US \$ unless otherwise noted.

The cost of the central XSEDE application, allocation, support, training, operation and security activities was simply the annual expenditure of federal funds through the XSEDE2 award [20]. Additionally, in a typical year, at least one system was added or phased out. The annual cost of each of the large CI systems associated with the XSEDE ecosystem was calculated as follows:

$$\text{Annual Cost} = \frac{\text{Months Available}}{12} * \frac{\text{Total Cost to NSF}}{\text{Total Months Available to US researchers}}$$

Months available were counted as months available for use by the national research community, not including pre-production (“early user”) availability. Several new CI resources were introduced in the last years of XSEDE that will remain available to the national US research community for some time after the project’s end. We estimated 60 months as the time these resources would be available to the national research community, consistent with past history and NSF grant award documents available online.

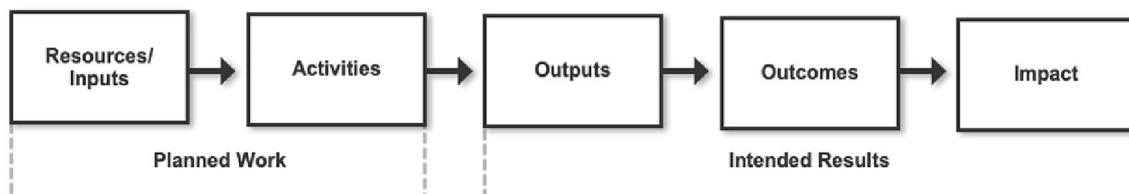


Fig. 1 Logic model of organizational processes

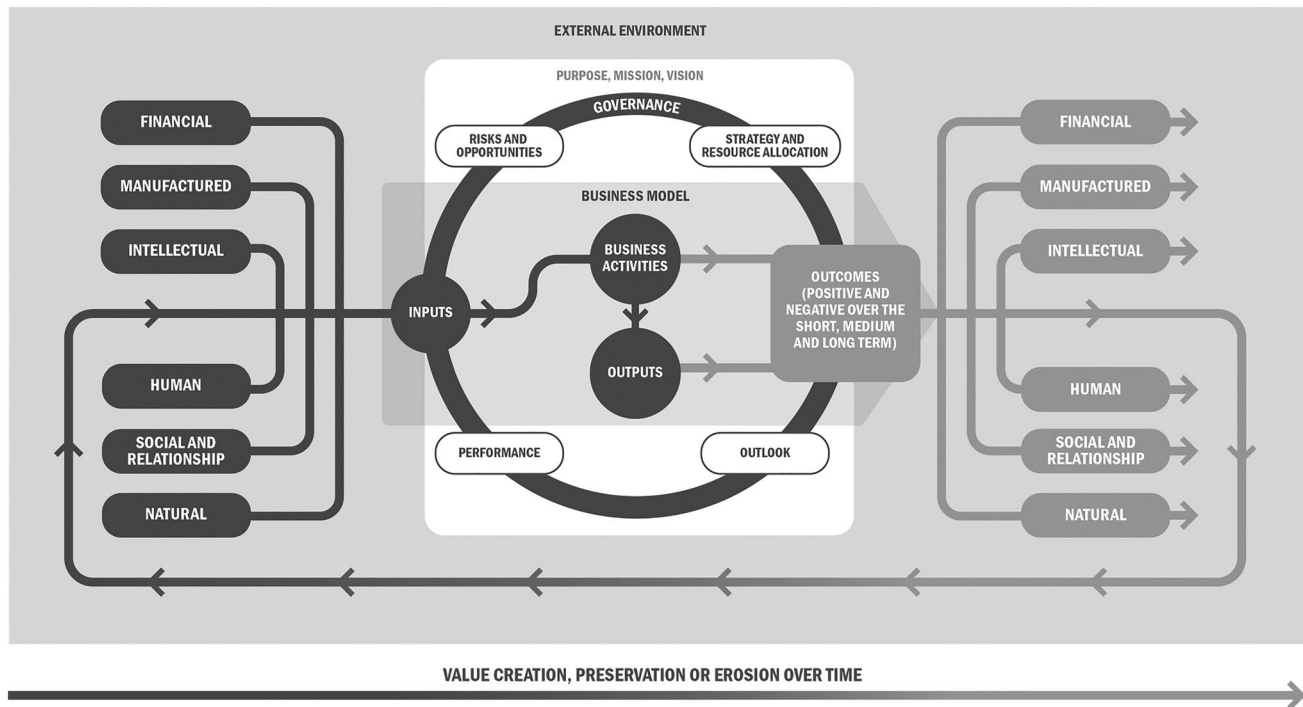


Fig. 2 Structure of <IR> Framework. (c) IFRS Foundation. Used with permission

To estimate the value of benefits resulting from the use of the XSEDE ecosystem, we identified every end product that we could measure, organized by the <IR> Framework categories, then identified a per-unit value for as many of the items as possible. These valuations and the logic behind them are presented in the results section. Data sources included XSEDE reports, the XSEDE Metrics on Demand (XDMoD) portal [21], the XSEDE Digital Object Repository [22], surveys of XSEDE users [5] and data from operational records. For some entities we have precise, incontrovertible numbers, such as the number of peer-reviewed papers published. For other categories, where benefits are harder to quantify, we use two estimates to provide a reasonable range of potential values. As a low estimate of value we took one extremely conservative estimate, and then we also added another estimate that we believe to be more realistic (the approach taken in our most recent work [8]). The lower estimate is designed to be so conservative as to be difficult for a critic to argue against—a useful point in some discussions.

There is always a gap between when the use of a CI resource is complete and when the value of end results is realized. In addition, the values of end products of different sorts are realized on different time scales. The timeframe for a paper to be published is very different from the timeline for basic chemistry research to translate into the availability of a new drug. Funding for the XSEDE ecosystem was relatively stable over time and we observed that outputs such as

publications with short lag times were also relatively stable over time [7]. Thus, it seemed reasonable to measure values generated based on the use of the XSEDE ecosystem during the same time window as that in which we measured costs. As a practical matter and in the interest of full disclosure, we had funding to collect and analyze data during the operation of XSEDE.

By multiplying the count of items recorded with the estimated value attributed to each item, we can approximate the total value of the end products generated through the utilization of CI facilities and services offered by the XSEDE ecosystem. The next difficulty is determining how much of that total value to attribute to XSEDE. The Appendix provides a mathematical representation of the information we seek—the value created as a result of the XSEDE ecosystem that would not have existed otherwise and a formula defining what we mean by “percentage credit” indicating the contribution of the XSEDE ecosystem to the creation of end products. Unfortunately there is no way to calculate these quantities directly. The issue of measurement and counterfactual arguments is addressed in the “Discussion.” There is no good way to simply calculate or reason what percentage of the credit is due to XSEDE for any particular end product created via the use of XSEDE ecosystem resources. Thus, lacking any direct way to assign an amount of credit, we work backwards. Starting from the data we have, we calculate the minimum percentage of credit that must be attributed to end products created via the use of XSEDE in



order for the  $ROI_{\text{proxy}}$  for NSF investments to be at least 1.0. The question then becomes whether or not the calculated minimum percentages of credit seem plausible.

## Results

### Costs

Table 1 shows the costs to the federal government of the XSEDE ecosystem for FY2017–FY2022. XSEDE fiscal years begin each September 1. The XSEDE ecosystem involved other systems, institutions, and people, but those were not generally federally funded. The few small systems that were funded by the federal government are not significant in terms of the total cost of the facilities nor their total capacity.

### Value of End Products

Values of end products estimated for FY2017–FY2022 are shown in Table 2. Tallies are aggregated across the six fiscal years of project operations analyzed in this paper. We note in this table end products for which the XSEDE ecosystem was solely responsible and end products for which credit is shared because end product creation involved utilization of XSEDE and the activities of other organizations or people.

Table 2 presents two different columns of end product value totals, reflecting different levels of conservatism in estimating the value of end products. There are particular challenges in drawing cause-and-effect relationships between research carried out with XSEDE and the creation of end product value related to saving human lives and

protecting the environment. Because of the challenges in establishing cause-and-effect relationships, different people may have different views on the level of conservatism appropriate for estimating the value of end products attributable to the XSEDE ecosystem in these areas in particular. Rather than take any one position, we present a range of potential values for end products enabled by XSEDE and then assess the impact of so doing on plausibility of a contention that ROI for investment in XSEDE is at least 1.0. The text below explains the rationale for the figures presented in this table, all organized within the categories identified in the <IR> Framework.

### Financial Value Created

The <IR> Framework defines financial value simply as value created in the form of money or other financial instruments. There is one clear way in which XSEDE has aided the creation of financial value, and that is in terms of grant funding received.

*Grant Funding Received.* One way to look at XSEDE's financial impact is from the standpoint of individual PIs who received allocations of XSEDE ecosystem resources. For PIs, grant funding is both an end product—an outcome of writing grant proposals—and a means for conducting research. Data from the XDMoD Portal [21] shows that researchers using the XSEDE ecosystem received more than \$4.5B in extramural funding between FYs 2017 and 2022. When PIs indicate that XSEDE was important in their ability to obtain funding, it seems correct to view that funding as a financial outcome of XSEDE. PIs who mentioned XSEDE in their proposals and received extramural funding consistently indicate that referencing XSEDE as a resource

**Table 1** US Federal Government Costs for XSEDE Ecosystem FY2017–FY2022

Expense	FY2017	FY2018	FY2019	FY2020	FY2021	FY2022	Totals
<i>XSEDE</i>	\$19,079,866	\$20,638,174	\$21,079,301	\$21,697,368	\$21,532,488	\$27,402,654	<b>\$131,429,851</b>
Gordon	\$2,383,212						<b>\$2,383,212</b>
Stampede	\$10,370,371						<b>\$10,370,371</b>
Comet	\$6,069,662	\$6,069,662	\$6,069,662	\$6,069,662	\$3,540,636		<b>\$27,819,282</b>
Wrangler	\$4,086,580	\$4,086,580	\$4,086,580	\$340,548			<b>\$12,600,289</b>
Jetstream	\$2,416,067	\$2,416,067	\$2,416,067	\$2,416,067	\$2,416,067	\$805,356	<b>\$12,885,692</b>
Bridges	\$4,179,033	\$4,179,033	\$4,179,033	\$4,179,033	\$2,437,769		<b>\$19,153,903</b>
Stampede2	\$4,908,577	\$11,780,585	\$11,780,585	\$11,780,585	\$11,780,585	\$11,780,585	<b>\$63,811,504</b>
Expanse					\$3,975,000	\$5,300,000	<b>\$9,274,999</b>
Bridges2					\$2,858,333	\$4,900,000	<b>\$7,758,333</b>
Anvil						\$4,478,975	<b>\$4,478,975</b>
Jetstream2						\$4,491,666	<b>\$4,491,666</b>
Delta						\$2,041,667	<b>\$2,041,667</b>
<b>Totals</b>	<b>\$53,493,369</b>	<b>\$49,170,102</b>	<b>\$49,611,229</b>	<b>\$6,483,264</b>	<b>\$48,540,879</b>	<b>\$61,200,902</b>	<b>\$308,499,744</b>

Bold text indicates totals

**Table 2** Estimated value of end products created using XSEDE ecosystem (values in US \$ )

Category of value	Tally	Multiplier	Conservative total value	Upper limit of total value	XSEDE credit
<b>Financial value</b>					
Grants received By PIs	\$4.5B	58%	\$2.6B	\$2.6B	Partial
<i>Financial value subtotal</i>			<i>\$2.6 B</i>	<i>\$2.6 B</i>	
<b>Manufactured products</b>					
No Direct Outputs	0	Unknown	\$0	\$0	
<b>Intellectual products</b>					
Peer-reviewed papers	11,233	\$1,475	\$16,568,675	\$16,568,675	Partial
Nobel prizes	1	\$1,106,100	\$1,106,100	\$1,106,100	Partial
Patents	27	\$56,525	\$1,526,175	\$1,526,175	Partial
Improving open-source software			\$3,250,000	\$3,250,000	Full
Software	709	Unknown			
Datasets published	865	Unknown			
Licenses	6	Unknown			
Organizational capital			\$13,000,000	\$13,000,000	Full
<i>Intellectual products subtotal</i>			<i>\$35,450,950</i>	<i>\$35,450,950</i>	
<b>Human capital</b>					
<b>Training with HPC</b>					
Baccalaureate degree recipients	1,144	10% x \$77,874	\$8,908,786	\$8,908,786	Full
Graduate degree recipients	4,858	10% x \$96,178	\$46,723,272	\$46,723,272	Full
Encouraging technology adoption		Unknown			Full
<i>Human capital subtotal</i>			<i>\$55,632,058</i>	<i>\$55,632,058</i>	
<b>Social and relationship capital</b>					
COVID-19 HPC Consortium	54 person-months	\$12,817/month	\$692,117	\$692,117	Full
Human lives saved	0; or 1,000–10,000	\$1,000,000	\$0	\$1B to $\geq$ \$10B	Partial
Cultural impact and joy of discovery		Unknown			Partial
<i>Social and relationship capital subtotal</i>			<i>\$692,117</i>	<i>\$1.1B to <math>\geq</math> \$10B</i>	Partial
<b>Natural resources capital</b>					
Preservation of environment			\$1B	> \$10B	Partial
<i>Natural resources capital subtotal</i>			<i>\$1B</i>	<i><math>\geq</math> \$10B</i>	Partial
<b>Subtotals</b>					
Full credit to XSEDE			<b>\$0.07B</b>	$\geq$ <b>\$0.07B</b>	
Partial credit to XSEDE			<b>\$3.6B</b>	$\geq$ <b>\$10B</b>	
<b>Grand total</b>			<b>\$3.7B</b>	<b>\$4.7B to <math>\geq</math> \$22.7B</b>	

Italics indicates subtotals. Bold indicates totals

for CI-intensive research aided competitiveness of their proposals for grant funding. This perception was based on the belief that referencing XSEDE added credibility to claims in the proposal that simulations or data analyses could be carried out successfully. Chityala et al. [23] report that 58% of PIs surveyed indicated that XSEDE was “important” or “very important” in obtaining grant funding; applying that percentage to the total extrapolated grant funding awarded to XSEDE PIs, we calculate a total of approximately \$2.6B as the financial value created for these researchers.

### Manufactured Value Created

The <IR> Framework defines “manufactured value” as the value of “manufactured physical objects.”

NSF funding for XSEDE never directly contributed to manufacturing research. We gave this category a financial value of \$0, although XSEDE-enabled research may someday result in manufactured goods with significant financial value.

## Intellectual Value Created

The <IR> Framework defines “intellectual capital” as the value of “organizational, knowledge-based intangibles, including: intellectual property, such as patents, copyrights, software, rights and licences [sic]” and “organizational capital such as tacit knowledge, systems, procedures and protocols.”

*Scientific and technical publications.* Peer-reviewed technical papers and reports in scientific journals are a primary intellectual product enabled by the XSEDE ecosystem. A publication’s value is a matter of significant debate. One value we can assign is the price that many (though far from all) authors are willing to pay journals to provide “open access” to their papers—in some sense the value that publishers and authors place on such papers. We have found open access fees ranging from \$750 [24] to \$2200 per paper (the publisher of the present journal). We multiplied the total number of peer-reviewed papers reported by PIs through the XSEDE-operated reporting system by \$1,475 (the midpoint of open access fees we found, which is less than the per paper value estimated in [25]) to calculate a total value for peer-reviewed papers.

*Nobel prizes.* As discussed in [26], a Nobel Prize award results in various kinds of value to prize winners and potentially others. During the period we are considering, one Nobel Prize was awarded for research that used XSEDE: the 2017 Physics Nobel Prize, recognizing three US researchers and LIGO (the Laser Interferometer Gravitational-Wave Observatory) for the discovery of gravitational waves [27]. XSEDE computational resources and the involvement of XSEDE staff were critical to this discovery [8]. A minimum value for this outcome is the financial value of a Nobel Prize: in 2017 it was 9,000,000 Swedish Kronor or \$1,106,100 at the conversion rate of 1 Kronor = US \$0.1229 on October 3, 2017, the day this prize was announced [28]).

*Patents.* Patents vary widely in value, but institutions should rationally only file for a patent if its expected value exceeds the filing cost. We use \$56,525, the estimated full cost of filing a patent application [29], as the minimum estimated value of a patent.

*Open-source software.* XSEDE produced software for purposes of operational management (one aspect of which is mentioned in a different category of value below). In general, XSEDE focused its software development efforts on improving preexisting open-source software [7]. XSEDE invested more than \$3.25M to harden open-source software, and the improved versions were released back to the community. (Financial information provided by the XSEDE budget office).

*Other specific intellectual products.* The survey described in [23] collected data on several other types of intellectual outputs, such as datasets published, software packages and

routines released, and licenses executed. We could not find general value estimates for these products but included the tallies as impetus for future analyses.

*Organizational capital.* The <IR> Framework’s intellectual capital category includes “organizational capital,” including “tacit knowledge, systems, procedures and protocols.” The NSF required that XSEDE prepare a series of transition documents to aid ACCESS, the project that succeeded XSEDE as the management and organizational structure for NSF-funded CI that was made generally available to the US research community [30]. We can attach a value to these materials and the operational software handed over to ACCESS based on the early experiences of the XSEDE project. XSEDE received very little in the way of procedure and protocol documentation from its predecessor when XSEDE assumed the role of supporting advanced CI resources on behalf of the NSF. In XSEDE’s first two years, we estimate that 25% of XSEDE’s overall effort was spent creating documentation for operational and technical processes that should have been captured in transition documents and associated software artifacts prepared by XSEDE’s predecessor. XSEDE’s annual budget was \$26M for 2012 and 2013, its first two years of operation [31]. This suggests a total value of \$13M for the knowledge captured and transferred by XSEDE to its successor. (These estimates are based on the experience of the co-authors who were directly involved in the XSEDE startup).

## Human Capital Developed

The <IR> Framework defines human capital as “people’s competencies, capabilities and experience, and their motivations to innovate.”

*Practical training in the use of HPC.* One area in which XSEDE helped to develop human capital was in facilitating bachelor’s, master’s, and PhD training and education. A new study surveyed large STEM-oriented firms to assess the value of HPC training [32]. It showed that employers offer about 10% more in salary to new hires with recent degrees in STEM technologies (excluding computer science (CS) majors, for whom the market is presently very distorted). For 2019, the most recent data available, the average starting salary for non-CS STEM bachelor’s degree recipients was \$77,874, and it was \$96,178 for graduate degrees [33]. The median number of days in which undergraduate students in the XSEDE ecosystem ran jobs on HPC systems was 29; for graduate students it was 45. We took the execution of 30 or more jobs as indicative of real hands-on competence. To estimate the value created by XSEDE in providing this training, we multiplied the number of students who ran jobs on 30 or more days by 10% of the average post-degree starting



salary, to represent the value that hiring firms place on the possession of practical HPC skills.

*Facilitating technology adoption.* Another area in which the XSEDE ecosystem clearly helped to develop human capital in the US was facilitating the adoption of new technologies. The development, support, and popularization of science gateways is a particularly striking example. Science gateways automate a set of complicated tasks to achieve an overarching goal, such as searching for drug candidates or predicting a tornado’s path. They utilize advanced computational systems, allowing users to benefit from their capabilities without needing to be supercomputing experts [34, 35]. XSEDE supported a total of 49 science gateways by funding staff to help create them, providing mechanisms for hosting and running computational analyses on resources within the XSEDE ecosystem. The number of individuals accessing XSEDE resources through science gateways first began to exceed the number using traditional command-line access in 2014 [36]. XSEDE began putting extra emphasis on supporting new science gateways and promoting the use of science gateways at the beginning of the XSEDE2 award—the time period considered in this study. Figure 3 shows that the added emphasis on science gateways created a step function increase in the use of science gateways during the first year of the XSEDE2 award. This significant and rapid increase was primarily attributed to the growing usage of two science gateways: the I-TASSER gateway, which supports protein structure and function prediction,

and the CIPRES gateway, which supports phylogenetic inference [37]. CIPRES seems to have built up its user base over time. I-TASSER was integrated with XSEDE ecosystem resources in October of 2016 [38]. Our inference, particularly regarding I-TASSER, is that integrating it with the XSEDE ecosystem significantly enhanced its ability to support a larger number of users and simultaneously made it more easily discoverable for users. The increase in use of this gateway, in particular, indicates that there was a significant number of scientists and students who had developed an interest in these topics and had not previously considered the XSEDE ecosystem to be accessible or beneficial to their needs. Also clear from Fig. 3 is that growth in the number of users of the XSEDE ecosystem from 2016 through 2022 is driven almost exclusively by growth in the number of people using science gateways. Surveys and interviews show that many people who require advanced CI resources found science gateways to be user-friendly, while they perceived command-line access as less straightforward [31]. Other ways in which science gateways deliver value to researchers and thus indirectly to the general public include speeding access to codes by making resources available without having to apply for allocation of resources through XSEDE and avoiding the need for users to install complicated suites of software on their own local CI facility (Michael Zentner, personal communication). We are not sure how to quantify these values, but they certainly seem significant, particularly when applied to several thousands of XSEDE science gateway users per year.

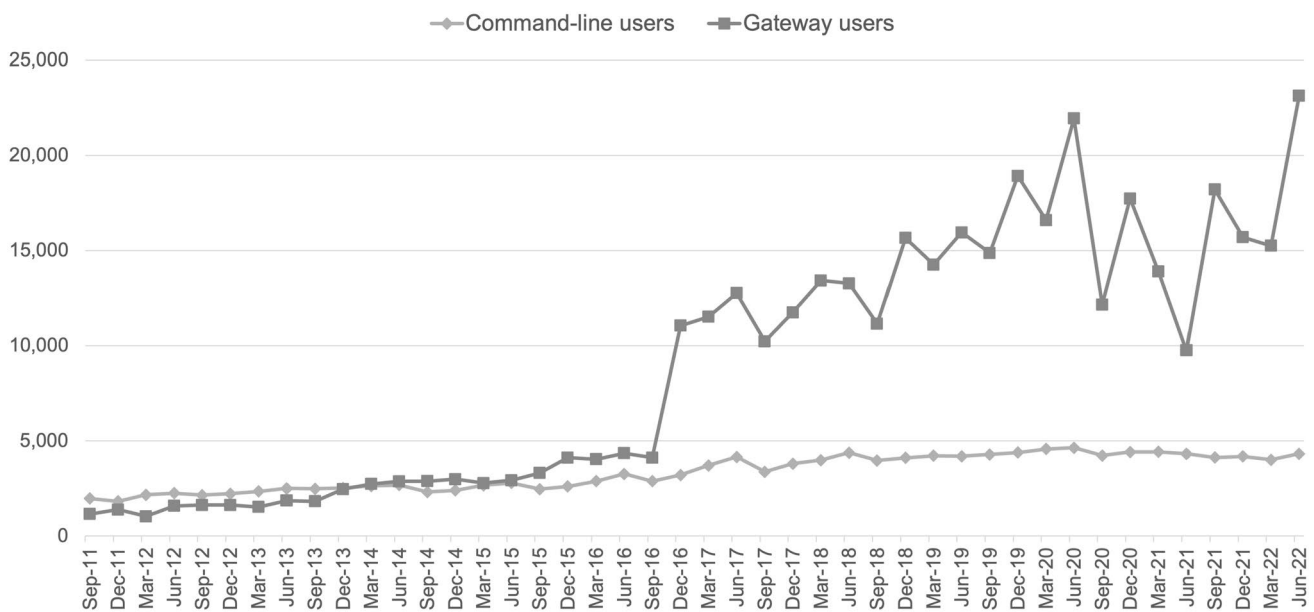


Fig. 3 XSEDE CI systems access via science gateways vs Unix command lines

## Social and Relationship Capital Developed

The <IR> Framework describes social and relationship capital as the “institutions and the relationships within and between communities, groups of stakeholders and other networks, and the ability to share information to enhance individual and collective well-being.”

The following sections describe XSEDE contributions to the creation of social and relationship capital in three forms: first, contributions to federal government emergency response during the early phases of the COVID-19 pandemic; second, estimates of lives saved where XSEDE played some contributing role at state and national levels; and third, qualitative comments about the value of improving quality of life.

*Supporting government emergency response.* Supporting the implementation and operation of the COVID-19 HPC consortium: in the early days of the COVID-19 pandemic, the federal government led the creation of the COVID-19 HPC Consortium (C19HPCC), a public/private partnership that provided HPC resources for research and development specifically targeted at fighting the COVID-19 pandemic. The C19HPCC resources ultimately comprised more than 300 PetaFLOPS of aggregate computing capability, including several resources that were part of the XSEDE ecosystem. The C19HPCC offered allocations of resources provided by more than 43 different consortium partners allocated on the basis of merit review, but without direct financial cost to researchers. An overview of the activities and accomplishments of the C19HPCC is online at [39].

The C19HPCC was initially discussed between IBM Inc. and US government leaders beginning the weekend of March 14–15, 2020, resulting in a decision that the White House Office of Science and Technology Policy would lead up the implementation and operation of the consortium. John Towns, XSEDE PI, was contacted to solicit XSEDE’s involvement on March 19. By Sunday, March 22, XSEDE had assembled and put into operation an allocation system for use by the C19HPCC, based on the pre-existing XSEDE allocation system. On that day, the president of the US announced the formation of the consortium. The first application was submitted on March 24 and computational work on the first approved project began on March 26 [40]. Thus the federal government benefited in terms of speed—an allocation system was in operation one week from initial discussion of the concept of the consortium—as well as financially, because it did not have to invest in creating such a system from scratch. It is not possible to put a value on the speed that the XSEDE ecosystem provided in helping the C19HPCC quickly become operational. But we can provide an estimate of the financial value that was delivered to the C19HPCC. The initial XSEDE allocation system took approximately 54 person-months to create. Co-author Hart

led this effort; this estimate is based on his first-hand experience. The reapplication of this software took place during the last six years of XSEDE, during which time the average annual cost of salary and benefits was \$153,803. This comes out to a financial value of \$692,117 for this software, which was provided to the C19HPCC at no cost. This is one of the cases in which XSEDE delivered value directly to the federal government.

*Supporting lifesaving research.* It is difficult but in some cases possible to quantify the number of human lives saved as a result of efforts to which XSEDE contributed. (How much credit for these shared efforts is appropriate to give to XSEDE is a different question, discussed later in this paper). In order to convert lives saved to a financial figure, one must assign a financial value to a human life. There are many estimates available for the financial value of a human life, ranging from \$325,000 for a baby to \$8.7M for a worker in an industrialized nation [41]. Here we take the very conservative figure of \$1M per human life (in line with a commonly observed wrongful death settlement or life insurance amount). Areas in which one can plausibly relate the saving of human lives to research supported by XSEDE are described below.

*Pre-clinical drug development.* XSEDE has supported a great deal of pre-clinical drug development. In the time period studied, the XSEDE ecosystem supported research resulting in a total of 1898 peer-reviewed publications related to drug development (see [5]). New drugs can of course sometimes be lifesaving and sometimes significantly improve quality of life. We don’t know how many of the publications related to drug development were part of research programs that identified a drug lead that went into clinical trials and then resulted in the availability of a new drug for patient use. The success rate for “non-lead” drug candidates (the kind of candidate for which pre-clinical screening trials would be done on XSEDE resources) is less than approximately 1% (cf. Supplementary Table 2 in [42]). Unfortunately, there seems to be no good way to estimate numerically the number of non-lead drug candidates that were successfully identified via research enabled by XSEDE without analyzing each one of nearly 2,000 publications. Still, it seems reasonable to believe that one or more drugs will eventually come to market as a result of research enabled by use of the XSEDE ecosystem. Creating a proper estimate is beyond the time and resources available in the execution of the current research.

*Severe weather prediction.* Developing and refining tornado and hurricane predictions was a particularly important use of the XSEDE ecosystem [7, 43]. Better tornado prediction software saves lives because people have more time to take shelter. For the years 2013–2022, the average annual number of tornado-caused deaths in the US was 45 [44]. For the prior 20 years, that average was 84. We can

infer that better modeling developed using the XSEDE ecosystem contributed to a reduction in tornado-caused deaths of 39 per year. This comes to a total of 234 lives saved over the 6-year time period analyzed. Some of these outcomes were time-lagged from the research that enabled them. Some work using the XSEDE ecosystem predates the most recent award, in some cases going back to XSEDE's predecessors. Work done using XSEDE during the time period studied can also be expected to contribute to saving lives in the future. With clear signs that tornado severity is rising, this work may save even more lives in the future.

*COVID-19 Nonpharmaceutical Intervention Policy Responses and Human Lives Saved.* The COVID-19 HPC consortium supported a wide variety of research relevant to COVID-19. We now know that in the short run, the use of these resources to enable drug development research, in particular, was not important in terms of the US response to the pandemic. XSEDE ecosystem resources, however, also supported a great deal of research on the genetic modification and creation of new variants of the virus over time, research designed to elucidate how the virus was transmitted, and research related to the effectiveness of nonpharmaceutical interventions (NPIs) (cf. [45]), all of which have some relevance to the early phases of the response to COVID-19. A bibliography compiled of publications resulting from use of C19HPCC shows that a total of 23 papers relevant to COVID-19 responses were published using XSEDE resources. Of these, 15 publications related to NPIs as tools for fighting the pandemic were confirmed by their authors or by the presence of acknowledgments in the paper itself to have been based on computations done using the resources of the XSEDE ecosystem [46].

Some of the research that resulted from the use of XSEDE ecosystem resources appears to have been significant and influential in understanding the spread of the virus indoors, the role of face masks, genetic changes in the virus, and modeling the pandemic. Distinguished Professor Madhav Marathe, head of the Biocomplexity Institute at the University of Virginia, and his colleagues indicated that they believe research enabled by the XSEDE ecosystem played an important role in setting policy related to government response to the pandemic at multiple levels. Professor Marathe stated,

Our work presented the first and only sustained use of HPC resources for supporting the CDC Scenario modeling hub. It started with XSEDE... This to me is the first demonstration of "real-time" planning and response for epidemics using HPC resources. No other group has been able to do this and... [XSEDE] support was critical for us to start and sustain this...

Use of HPC allowed us to use and demonstrate the value of a qualitatively different class of models, those based on social networks... This ability is a game changer.

Marathe and his colleague Machi confirmed that their center's work influenced pandemic responses at multiple levels of government: in the Commonwealth of Virginia through work with the Virginia Department of Health; at military facilities through work with the Department of Defense and Defense Threat Reduction Agency; and throughout the US through work with the Centers for Disease Control and Prevention (CDC), the CDC Scenario Modeling Hub and the National Science Foundation (Madhav Marathe and Dustin Machi, personal communication).

The University of Virginia Biocomplexity Institute's use of XSEDE ecosystem resources via the C19HPCC and during the time period studied in this paper resulted in a total of 8 peer-reviewed publications (see [46]), including the following three directly related to policy steps involving NPIs: [47–49]. In the wake of the pandemic research, Professor Marathe was honored by the Indian Institute of Technology, his alma mater, with its Distinguished Alumni Award [50] and Marathe's institute received a letter of acknowledgment from the Commonwealth of Virginia.

How to estimate the number of lives saved at the national level is a challenge, but there are statistics that give us ways to approach that issue at the level of the Commonwealth of Virginia. During 2020—the time before vaccines were generally available—5821 people died in Virginia of COVID-19 [51]. Virginia had one of the lower per capita rates of COVID-19 mortality during 2020: 56.3 deaths per 100,000 residents, the 10th best such rate in the US [52]. There are at least three ways one could estimate lives saved in Virginia, or the US as a whole, as a result of NPI policies during 2020, the year before vaccines were available (all calculations detailed in [5], with CDC mortality data and Census population data as the data sources):

- *Lives saved in Virginia estimated by comparison with national averages.* The US national average for COVID-19 related deaths that year was 105.6 per 100,000. Based on a state population of 8,631,373, there were 3293 fewer deaths due to COVID-19 in Virginia in 2020 than would have been expected on the basis of national averages.
- *Lives saved in Virginia estimated by comparison with death rates in adjoining states.* Virginia shares borders with Maryland, West Virginia, Kentucky, Tennessee, and North Carolina. The average mortality rate in these five states due to COVID-19 in 2020 was 88.3 per 100,000. Based on this, there were 1803 fewer deaths due to COVID-19 in Virginia in 2020 than would have been expected on the basis of averages from bordering states.

- *National mortality statistics.* Figures from the CDC showed that a total of 1,146,351 people had died of COVID-19 as of December 22, 2023 [53], the first date on which vaccines became widely available in the US. NPIs were crucial in reducing mortality rates in the US prior to this date. One particularly careful estimate shows that the net number of lives saved through policies promoting use of NPIs was between 808,428 and 1,466,095 [54].

Many factors contributed to mitigation of the impact of the COVID-19 pandemic at the state and national level. We have first-hand researcher statements that confirm that research supported by XSEDE ecosystem resources contributed to policies at the state level in Virginia and at the national level, via analyses and policy recommendations made by the CDC and DOD.

*Assigning a value to lives saved where some credit is attributable to the use of XSEDE ecosystem resources.* In all of the above cases, it is a challenge to draw a specific causal link between research enabled by the XSEDE ecosystem and human lives saved. Thus one very conservative approach is to assume the number of lives saved to be 0. This is done for the most conservative estimates of end product value created by XSEDE in Table 2. However, this seems very pessimistic—unduly pessimistic, we believe, given the statistical analyses we present, as well as researcher testimonials.

A less conservative estimate of human lives saved as a result of research supported by XSEDE during the six years studied in this paper is to aggregate the following:

- We have reasonable belief, but no way to quantify, that XSEDE contributed to saving some lives due to support of early-phase drug candidate screening.
- There is credible evidence that XSEDE and its predecessors contributed to a reduction of 39 lives per year, or a total of 234 lives, due to a tornado prediction study.
- There is clear evidence that research supported by XSEDE contributed to saving an estimated 1803 lives in the Commonwealth of Virginia alone due to policies related to NPIs.
- At the national level, take the midpoint of the estimated range of lives saved as a result of NPI-related policies in the US. We have accounts from researchers that state that XSEDE-supported research contributed to national policy setting. If XSEDE played some role in just one tenth of one percent of the midpoint of the range of human lives estimated to have been saved as a result of NPIs in the early phases of the pandemic, that would imply that XSEDE merits some credit for saving 1137 lives.

One can easily imagine XSEDE-enabled efforts having had more impact than these numbers indicate, given that XSEDE-supported research served as input for policy-making at the national level by the CDC and DOD.

Considering drug discovery, severe weather prediction, and involvement in the early phases of US response to the COVID-19 pandemic, it seems quite plausible that the use of XSEDE system resources contributed to efforts that resulted in saving somewhere between 1000 and 10,000 lives or more during the time period studied. This range for the number of lives saved is the basis for the upper estimates of end product value enabled by XSEDE in Table 2. If one assumes that XSEDE played some partial role in saving just 1000 lives, that implies a financial value of \$1B. Extrapolating more generously from state-level and national data and estimating some role for XSEDE in saving 10,000 lives or more implies that XSEDE was responsible for some portion of a financial value of \$10B or more.

Another view of the financial value of XSEDE contributions to US pandemic response prior to the availability of vaccines can be obtained by considering the very careful modeling analysis by Chen et al. This analysis, itself based on the use of XSEDE resources, shows that prior to the availability of vaccines there was essentially a trade-off between economic losses due to restriction of the economy as a result of lockdown policies and losses in terms of human deaths [55]. Chen et al. showed that decreasing one increased the other. Another view of the cost of the pandemic is a total long-term estimate of \$16T [56]. The range of estimates presented here as the economic value of XSEDE activities are a tiny fraction of this amount, lending plausibility to the idea that the national reach of XSEDE-based research can be measured in the billions of dollars.

It should be noted that COVID-19-related research carried out using XSEDE resources via the C19HPCC also has the potential to support future pandemic responses. Professor Marco Giometto stated that, thanks to research done on XSEDE resources allocated through the C19HPCC, “we were recently able to secure a large external grant from the NSF, whose findings will hopefully have an important impact on how we respond to pandemics” (personal communication; grant information at [57]).

*Other social capital benefits that seem impossible to quantify.* There are other kinds of social capital that are important to recognize even if impossible to quantify. These include the effects of outreach and educational activities, cultural impact, the pure joy of discovery and the wonder of gaining a deeper understanding of our universe [58]. Through outreach events, staff in the XSEDE ecosystem reached thousands of young



students and laypeople in dozens of states and the territories of Puerto Rico and Guam [7]. As [58] points out, large research infrastructure can have economic benefits through cultural impacts. The XSEDE ecosystem was funded to support science but also supported humanities research and the creative arts [7]. XSEDE supported the “joy of discovery” in many ways, including any number of scientific discoveries and visualization of those discoveries, such as an extremely popular visualization of black holes [59].

### Natural Capital Developed

The <IR> Framework defines “natural capital” as “all renewable and non-renewable environmental resources and processes that provide goods or services.” It includes “air, water, land, minerals and forests” and “biodiversity and ecosystem health.” This includes “continued availability of natural resources,” including mitigation of natural disasters and the effects of global climate change.

The XSEDE ecosystem enabled a large number of important research discoveries related to natural capital, including wildfire modeling and general responses to global climate change.

*Wildfire modeling.* The XSEDE ecosystem contributed significantly to wildfire modeling (see [60] for a recent example). Damage by wildfires is increasing; its annual value has been estimated at \$63B to \$285B [61]. Unfortunately, we lack sufficient information to assign a financial value to XSEDE’s contributions.

*Amelioration of global climate change.* The total cost to the US of climate change can be estimated at \$3,604B per year, based on the Gross Domestic Product and data in [62]. Taking a much more constrained view, the minimal and most conservative value of XSEDE-enabled investment in the preservation or improvement of environmental quality is \$1B. This is the value of investment by the State of California resulting from research enabled by the use of XSEDE ecosystem resources [63]. One can easily imagine the investment in or outcomes from efforts to preserve our natural environment totaling \$10B or more, assuming that the effects of XSEDE-enabled research lessen the cost of global climate change to the US economy by even a small amount.

### Other Financial Value Of Note

The XSEDE ecosystem has generated or facilitated the generation of other valuable outcomes that extend beyond end products. In particular, there is anecdotal evidence of extra effort contributed by individuals and institutions to support the overall XSEDE ecosystem, often (based on anecdotal evidence heard by the authors of this paper) in the belief

that so doing was in the best interests of the colleagues of the individuals, or the researchers at the institutions, that invested such extra effort. In general it is impossible to quantify the value of these efforts. There is one case, however, where we are able to estimate the value of extra efforts contributed to aid the accomplishment of the overall XSEDE ecosystem: the efforts of a group of staff known as Campus Champions. Campus Champions were a formalized group of on-campus staff who supported the use of the XSEDE ecosystem [7]. On average, 567 Campus Champions were active at any given point during the time period studied. Using a very conservative estimate that they spent 5% of their time engaging in XSEDE activities, that would mean that they gave 170 total FTE-years to the XSEDE ecosystem over the 6-year period studied here. Campus Champions were sometimes not quite as expert as XSEDE staff, so we use a conservative estimate of 2/3 of the XSEDE staff cost; thus, they contributed an estimated \$17,441,611 worth of effort [5], typically funded by the institutions for which the Campus Champions worked.

These financial benefits, while notable, are not financial end products per se. They are inherently included in calculations of  $ROI_{\text{proxy}}$  by virtue of *not* being represented in the denominator, since they reflect costs not borne by the federal government. These financial benefits are a direct effect of federal funding for the XSEDE ecosystem. One function of this funding is to create an initiative to which other individuals and institutions choose to contribute.

### Calculating $ROI_{\text{proxy}}$

How do we go about calculating a proxy for ROI for the federal government’s investment in XSEDE in terms of the value of end products created through the use of its ecosystem? This cannot be estimated using counterfactuals, such as “What would have happened if XSEDE had never existed?” We can make educated guesses about what would have happened if XSEDE had not existed and there were no other options. But other options did exist. If XSEDE were not present, certain research would have been carried out just as effectively using alternative approaches, while some research would have been conducted with reduced efficiency, and certain research projects would not have been pursued at all. But we can take a different approach. Assuming that the XSEDE ecosystem gets some credit for end products where it was among many entities that contributed to an end product, we can calculate the average percentage credit that must be attributed to the XSEDE ecosystem in order that  $ROI_{\text{proxy}}$  be at least 1.0, working from the formula below.



**Table 3** Minimum credit due to XSEDE ecosystem for  $ROI_{\text{proxy}}$  to be  $\geq 1.0$ 

Scenario	Total value estimated for end products of XSEDE	Estimated value of end products fully attributable to XSEDE	Estimated value of end products partially attributable to XSEDE	Cost of XSEDE to US Federal Government	Minimum percentage attributable to XSEDE for $ROI_{\text{proxy}} \geq 1$ (%)
Most conservative	\$3,701,775,125	\$72,574,175	\$3,629,200,950	\$308,499,744	6.50
Conservative, add lower estimate of lives saved	\$4,701,775,125	\$72,574,175	\$4,629,200,950	\$308,499,744	5.10
Higher estimates of lives saved and natural resources	\$12,711,775,125	\$72,574,175	\$22,629,200,950	\$308,499,744	1.04

$$ROI_{\text{proxy}} = \frac{(\text{Value of products due to XSEDE alone}) + (\text{Value of shared credit end products}) * (\% \text{ XSEDE Credit})}{\text{Costs of XSEDE ecosystem to US federal government}}$$

It is important to remember that these calculations do not suggest that an XSEDE staff member miraculously saved a child from oncoming traffic or that someone would have suffered a fatal heart attack had XSEDE not contributed to the creation of a black hole visualization. For the majority of the end products enabled by XSEDE, XSEDE is one of many factors that collectively enabled the creation of an end product (like lives saved through public policies set during the early days of the COVID-19 pandemic). The question at hand is: What percentage of the overall credit should XSEDE get? Table 3 shows the minimum percentage of credit that must be assigned to XSEDE for its contribution to the creation of end products, based on the equation above, in order for  $ROI_{\text{proxy}}$  to be at least 1, for three different levels of value assigned to the XSEDE ecosystem's contributions to saving human lives.

## Discussion

We can now make a plausible argument, using exclusively financial terms, that the return on—the value generated by—XSEDE for the US and its citizens is greater than its cost to the federal government. If, as shown in Table 3, the XSEDE ecosystem is assigned somewhere between 1.04 and 6.50% of the value of the end products to which it was one of multiple contributors, using the estimates presented here for the overall end product value created through use of the XSEDE ecosystem, then  $ROI_{\text{proxy}}$  for the federal government investment in XSEDE is at least 1. How reasonable is it to attribute 1.04–6.50% of the credit to XSEDE for value partially created using the XSEDE ecosystem?

Consider that for every year covered by this study, XSEDE users rated its importance to their research or

education as higher than 4 (on a scale where “1” was *not at all important* and “5” was *extremely important*). Among PIs engaged in federally funded research, 63% said that they could not have completed their research without the XSEDE ecosystem's resources [5, 7]. This seems to make our relatively modest attributions of credit to the XSEDE ecosystem seem plausible. For  $ROI_{\text{proxy}}$  to be at least 1 using the most conservative estimates of the value created by using the XSEDE ecosystem, the average percentage credit attributed to the XSEDE ecosystem must be at least 6.5%. That figure is based on an assumption that XSEDE played no role in saving any human lives, which seems far too pessimistic a view of XSEDE, given the accounts of researchers who used XSEDE and statistics about death rates in the US. Using the higher—and we believe more reasonable—estimates of the value created implies that the average credit attributed to XSEDE needs to be somewhere between 1.04 and 5.10% in order for  $ROI_{\text{proxy}}$  to be at least 1. These percentages would be even lower were we able to attach specific financial values to several items described, but not quantified, here. We have shown before that XSEDE processes were cost-effective. We now can answer whether the federal government's investment in XSEDE was worth it in strictly financial terms. Our answer is: yes—plausibly, if not conclusively.

For most outcomes discussed here it is straightforward to conceptualize the role of XSEDE in creating outcomes even when there are many other contributors to an outcome and even if it is not always straightforward to estimate how much credit XSEDE should be allocated for the creation of that outcome. Scientific papers and Nobel Prizes that depend on computational work or simulations done using the XSEDE ecosystem are examples. In the case of benefits to preservation of the natural world, we have a specific investment in one state to point to, based on research done using the XSEDE ecosystem. In other cases it is harder to trace the

cause-and-effect link from XSEDE to particular outcomes. It takes more years than XSEDE was in existence (in general) to complete the path from initial computational research on drug candidates to clinical availability of new drugs. As regards the COVID-19 pandemic, it is difficult even now to look back and tease apart what efforts had a causal effect in saving lives and which were pursuits of either dead ends or new discoveries that may be useful in the future but did not play a role in the responses to COVID-19. For this reason, we have taken the approach of calculating the minimum percentage credit required for return on investment in XSEDE to be at least 1.0 at three different levels of conservatism in assigning credit for outcomes to XSEDE, including one estimate so conservative that it assigns a value of 0 to the number of human lives saved as a result of research and development done using XSEDE. This seems to be an unreasonably pessimistic view of the value of the XSEDE ecosystem in this regard, considering the views, already described, of people who were directly involved in COVID-19 responses and the statistical analyses of things like deaths due to tornadoes and COVID-19 mortality rates. So if we strike the most conservative tallying from consideration, federal government return on investment is greater than 1 for a percentage of credit attributed to XSEDE ranging from 5.1% (if we assign a value to environmental value created and lives saved of \$1B each) to 1.04% (if we assign a value of \$10B to each of these categories), and we end up with even lower percentages if we are even more generous in estimating total value that is at least partially attributable to XSEDE.

Yet another and different worry about the value of XSEDE in ameliorating the effects of the COVID-19 pandemic is the thought that the pandemic was a once-in-a-generation event and that the value conferred by XSEDE in the case of the COVID-19 pandemic is thus unlikely to be replicated in the near future. However, rare events of all sorts occur from time to time, from pandemics to major weather events to stock market crashes. During the first five years of the XSEDE ecosystem (and thus prior to the time period analyzed in this paper), XSEDE staff and XSEDE resources aided O'Hara and her coworkers in demonstrating that computerized odd-lots trading had the potential to destabilize stock markets [64, 65]. This finding resulted in changes to reporting regulations for the NASDAQ and New York Stock Exchange [66]. XSEDE played a critical role in supporting this research (Mao Ye, personal communication). What is the cost of a stock market crash? That is of course hard to know; the massive crash of 2008 to 2010 cost the US economy trillions of dollars [67]. Did the work of O'Hara and coworkers actually prevent a stock market crash? That's impossible to know, but it is straightforward to see that regulatory changes enabled by XSEDE made the stock market safer. Indeed, a blueprint for a "National Strategic Computing Reserve" was published in a report

issued by the National Science and Technology Council [68]. Such a reserve would be organized to be very rapidly available in order to support US responses to national emergencies. Among the resources that most closely match what is called for in this blueprint are XSEDE and its successor, the ACCESS program. Thus, it seems appropriate to attach some value to the potential for XSEDE and its successors to serve as a resource in emergency situations, much as insurance policies have a value even if no claims are ever filed against them. The exact value that should be attributed to XSEDE in terms of saving lives and emergency responses is very difficult to quantify, however, and this is a potential area for useful future research.

Given the relative novelty of the use of the <IR> Framework in the economic analysis of scientific endeavors, some comparison with other methodologies previously used seems warranted. A number of other frameworks have been used in studies of ROI in science. Each has strengths and weaknesses.

A framework developed in Florio and Sirtori's 2014 paper [58] is complementary to the <IR> Framework. In particular, it includes an interesting discussion of the kinds of benefits created by investment in science, including, for example, the financial value of the cultural activities built up around NASA's Kennedy Space Center. This framework was used in a cost-benefit analysis of the CERN Large Hadron Collider [69] that suggests a high likelihood that the return on investment for that facility will be greater than 1. An issue with this analysis is that it is based on the use of "shadow pricing" to estimate values. Shadow prices are calculated values for entities that are neither bought nor sold [70]. One kind of shadow pricing is called contingent pricing. This involves customer interviews (or surveys), asking their assessment of the values of various entities in hypothetical situations. This approach is widely used in assigning value to natural environments and environmental preservation (for example, see a special issue of *Ecosystems* [71]). The challenge for research subjects in such studies is predicting what they would do in hypothetical scenarios. This can be a risk to the precision of this approach [72].

ROI in research by the government of the United Kingdom has been estimated at 1.00–1.25, using a sophisticated statistical modeling framework [73]. This is a very interesting approach but depends on a great deal of knowledge about economic activity within the UK. It is possible to apply this approach in other situations, but doing so requires that same great degree of information about economic activity in any new setting where this approach might be applied.

Within the US, one model called Regional Input–Output Modeling System (RIMS II) is an approach based on "regional multipliers" and commonly used to estimate the economic benefits of government investments [74]. The Regional Input–Output Modeling System uses empirically

derived multipliers statistically estimated by the US Bureau of Economic Analysis to relate investment in a market sector to direct and indirect economic impacts. This sort of input–output modeling approach has been generalized to a large number of countries in the IMPLAN (short for “impact analysis for planning”) system [75]. These approaches have the benefit of being widely accepted, in part because they are based on actual economic data. They have the disadvantage that the multipliers and models are based on relatively broad areas of economic activity. For example, research at an institution of higher education in the US might fall into the general categories of “junior colleges, colleges, universities, and professional schools” or “professional, scientific, and technical services” [74]. Because of the generality of the categories used, they may not be as precise at estimating end product value as actually cataloging value, in the way that we have attempted to do here.

We differ with one of the methods in a paper that assesses the value of the European Bioinformatics Institute (EBI) [76]. One assessment of the value of EBI assumes that when a researcher indicates that their research “could not have been completed without this facility,” then the total value of the resulting research may be credited to the EBI. This sort of counterfactual reasoning seems off base. It ignores the cleverness and dedication of scientists who are in the business of doing things not easily done. If XSEDE or EBI had not existed, many scientists would certainly have been dedicated enough to find or create an equivalent alternate resource and perform their research as well, or nearly as well, without the facility that they actually used. If one wants to investigate ROI in a CI facility, we believe it is essential to take into account the fact that credit for the creation of end products is shared across many entities. Failure to do so overlooks philosophical analyses of counterfactual reasoning and our understanding of the dedication and cleverness of scientists.

Florio and Sirtori’s 2014 paper makes a critical point about investments in large scientific infrastructure projects. One cannot really know in advance what is being paid for when one invests in scientific research or in facilities supporting scientific research: the end product is fundamentally discovery and it is impossible to tell in advance what will be discovered [58]. This means that any prospective attempt to estimate ROI will have inherent limitations, as will any approach based on economic models that are themselves based on economic averages.

To the best of our knowledge, the <IR> Framework has never before been used to analyze CI facilities. Neither our implementation of the <IR> Framework nor the other approaches mentioned above is a perfect reflection of reality. However, of the available frameworks, we find it to have the most comprehensive categorization of types of benefits. Using it has spurred our thinking and expanded our analysis.

It also has the benefit of being backed by an international consortium of accounting experts and used across multiple economic sectors, creating a basis for potential comparisons. Compared to the other methodologies mentioned, our implementation of the <IR> Framework is distinctive and particularly useful because it involves directly measuring outcomes and estimating value from market bases. We have had reasonable success in finding financial values for many activities that XSEDE counted and reported. Much of that success resulted from consulting the applied economics, actuarial, and research policy literature. Some of it came from being willing to rely on credible, though not peer-reviewed, analyses available online. Comparable market prices and actuarial estimates of value provide a way to measure and convey value that is sensible and straightforward to explain. This means that the valuations we have used, in some cases at least, are based on reference values in the actual marketplace of the US, and these are more intuitive (and perhaps to some readers, thus more credible) than the use of shadow pricing. Similarly, we note that this approach is rooted in specific measurements of the outcomes enabled by XSEDE, rather than being based on general economic models, as is the case for RIMS II or IMPLAN.

Prior analyses of the services operated under funding for the original XSEDE and follow-on XSEDE2 awards yielded an  $ROI_{\text{proxy}}$  between 2 and 3. That is, it would have cost two to three times more to purchase services on the commercial market than to obtain them via NSF grant awards. Several analyses, including [18], show that the NSF grant awards to create CI facilities are markedly less costly than it would be to purchase services from commercial cloud providers. In other words, the ROI to the federal government for funding the creation and operation of facilities through grant awards as compared to purchasing from commercial cloud providers is much greater than 1. Combining these results with the analysis presented here, we can say that the XSEDE ecosystem provided services to the research community in a financially efficient manner and that the value of the results delivered exceeded the cost of operating the XSEDE ecosystem. In summary, both for operating XSEDE and creating end products,  $ROI_{\text{proxy}}$  to the federal government was greater than 1. In the case of the value of end products,  $ROI_{\text{proxy}}$  may ultimately be much greater than 1 over time, after US society has fully reaped the benefits of the research and development enabled by the XSEDE ecosystem.

Our previous work and the study presented here suggest the possibility of using two complementary approaches, one that can be used straightforwardly in the short term and one that can likely only be effectively used in the long term. Our earlier work demonstrating methods for evaluating the financial efficiency of the delivery of CI resources and services can be used in the short run to determine whether a CI service is offering good value

to the organizations funding it. This method can be used over time periods as short as a year and compares the cost of services delivered relative to the general market costs. This approach is also of use in longer-term analyses of the financial effectiveness of operational activities. For a more comprehensive analysis, and perhaps only effective in the long term, the approach demonstrated here and rooted in the use of the <IR> Framework can be used to assess the value of the end products created through use of a CI facility. Here we offer a longer-term approach to the value of a particular project as measured after the fact, in terms of end products enabled and created. These two approaches complement each other nicely. Our sense is that the approach taken here to assessing the value of end products may work better for small projects than for large ones. For smaller CI projects, it may be easier to assess the value of specific end products than for a very large project. In the case of the XSEDE ecosystem, we simply didn't have the funding or time required to ascertain in detail the value of specific end products created. Our dual approaches could also be applied to projects that exclusively involve software creation. For example, there are good metrics for the cost of software development from the commercial sector. One could evaluate the financial effectiveness of a software project by comparing its actual code production costs with industry benchmarks. Furthermore, one could use the approach presented here to put a value on the end products created. For example, the Science Gateways Community Institute (SGCI) [34] was responsible for the software behind the XSEDE-supported science gateways. SGCI could justifiably conclude from the data presented here that the value of the software it produced enabled access to XSEDE resources by roughly half of the XSEDE ecosystem's users, which could in turn justify financial arguments about  $ROI_{\text{proxy}}$  for investment in SGCI. This example also shows why reasonable estimates for the percentage of credit for the creation of end products given to any particular CI facility product must be relatively low, since many different facilities and services can fairly lay claim to having contributed to XSEDE's success. In this case, SGCI and the XSEDE ecosystem share credit for the end products created via the use of science gateways, since SGCI was responsible for initial software development and XSEDE adapted and refined this software.

Potential threats to the validity of the analyses presented here include:

- *Use of a "point in time" analysis rather than a longitudinal study.* A longitudinal study would enable detailed examination of startup costs and a longer-term analysis of the value of XSEDE-enabled end products. Startup costs were likely modest. XSEDE was one in an ongoing series of NSF-funded programs to provide

advanced CI services, and users are attuned to periodic changes in how services were delivered. However, we have certainly underestimated the long-term value of some of the research enabled by the XSEDE ecosystem that might be revealed in a longitudinal study.

- *Accuracy of the valuations assigned to end products.* Recognizing this issue, we have been extremely conservative in our estimations.
- *The credit allocated to the XSEDE ecosystem for its contributions where credit is shared.* We believe our argument here makes a credible beginning on this point.

There is another issue that merits mention but is not a threat to the validity of the analyses we present. Some of the costs of the XSEDE ecosystem were paid by entities other than the federal government, as mentioned. Indeed, one function of federal investment in XSEDE was to create an initiative that attracted support from other sources. If one performs the same sort of analysis that we have carried out here for the government's investment in XSEDE, but considers total investment in the XSEDE ecosystem overall, the final conclusion about delivery of value relative to cost is unchanged.

Opportunities for extending and improving the sort of analysis presented here in future analyses of CI facilities include:

- *Evaluation of the value of specific end products rather than averaging over categories, including longitudinal studies.* Where possible, assessing the value of each specific end product would lead to more robust valuations than the approach taken here. This would be particularly valuable, for example, in identifying the role of CI facilities in research leading to new medical treatments. Longitudinal studies of pre-clinical drug discovery research through the entire drug development (or policy development) pipeline would help tie basic research more definitively to eventual benefits to human health and the saving of lives.
- *More focus on the value of technology adoption and data products.* More focus on these areas would be valuable, using methods brought to bear in studies of the European Bioinformatics Institute (EBI) [76] and the Protein Data Bank (PDB) [77].
- *Use of contingent pricing to assess natural capital.* Contingent pricing, based on interviews and surveys, is the most widespread method currently used to assess value of the natural environment. Using this approach to evaluate natural capital would be more consistent with other work in environmental sciences assessment.

This paper includes three primary contributions. The first is the introduction of the <IR> Framework to the CI community and demonstrating its utility in the assessment of a



CI facility. The second is the demonstration of a credible first attempt at analyzing the value of end products created through the use of a CI facility. The third is the presentation of an argument that makes plausible the assertion that the value delivered to the US citizenry and economy by XSEDE is considerably greater than its cost to the US government.

We believe that this paper adds to the methodology of ROI studies for research and research facilities in several ways. It continues and reinforces our earlier emphasis on the concept of a proxy for ROI (typographically represented as  $ROI_{\text{proxy}}$ ) and on the use of comparable market values for analyzing the financial value of end products created. We also believe that our focus on the issue of partial credit for the contributions of facilities to the creation of end products is important. The contributing role of CI facilities as part of a community that collectively creates value is emphasized in [78]: “Cyberinfrastructure departs from the classic big sciences... which were goal directed large-investment collaborations of scientists and engineers (as with the Manhattan Project, the Human Genome Project, or the Large Hadron Collider). Instead, CI seeks to support research generally by identifying vast interdisciplinary swaths that could benefit from data and resource sharing, knowledge transfers, and support for collaboration across geographical, but also institutional and organizational divides.” This implies that any attempt to assess the value of a CI facility must deal with the issue of assessing how much credit to attribute to a CI facility when credit is almost always shared among multiple individuals, facilities, and organizations.

## Conclusions

We have made what we believe to be a reasonable argument showing that the US federal government investment in the XSEDE ecosystem results in a ROI greater than 1. The value returned to the US as a whole, and to US taxpayers in particular, exceeds the cost to taxpayers of funding the XSEDE ecosystem during the time period studied; this is credibly and plausibly, if not definitively, proven. As far as we know, this study is the first within the US to attempt to compare the *financial* value of the public good created by federal investment in a CI resource with the cost of that resource. We believe that the approach presented here holds promise and may be reused and refined in the future to aid the research community’s overall ability to evaluate and communicate the value of its work.

The modular and comprehensive nature of the <IR> Framework has greatly facilitated our analysis of this problem. Our prior work on the financial effectiveness of service

delivery and the work presented here offer two complementary approaches to analyzing the ROI for a CI facility: one that is applicable on a year-by-year basis for assessing activities during operation of the facility and one that is applicable in a summative fashion to assess the end products enabled. These approaches, while so far used mostly to analyze large hardware-oriented facilities, also seem applicable to software projects and smaller facilities.

We hope that this study will serve as a starting point for us and others to build on, as was the case for our 2015 paper on the financial effectiveness of XSEDE operations [79]. We can see how far our community has come in analyzing ROI for CI by comparing the papers in the current special issue of Springer Nature *Computer Science* with the slides from a panel discussion at the 2014 IEEE/ACM Supercomputing Conference, summarizing the state of affairs at that time: “the topic is hot, but solution is elusive” [80]. The work we show here offers tools that can be used in the long run to analyze, and in the short run to effectively convey, the financial value of investment in CI facilities supporting research. As pointed out in [81], no single number can adequately capture the return on a public investment in research. A multifaceted approach to assessing the value of CI is needed to robustly make the case for continuing such investments in the future.

## Appendix: Apportioning Credit for Creating Value

The value created over time by any organization or endeavor is a function of many factors, including levels of investment, public acceptance of innovations, disease, climate, war, etc. It is possible, at least in theory, to mathematically relate these factors with the value created, as a combination of functions of investments and external factors other than these investments. Let:

- $f_i$  represent a series of functions that relate value created to a number of factors including: government investment (represented by  $I$ ) which is time-dependent, and other external factors (represented by  $\epsilon_i$ );
- $e_i$  represent other functions that create or destroy value, dependent on time and other external factors  $\delta_i$

Investments and external factors may interact with each other in a number of different ways, as may external events. For the sake of a reasonable degree of simplicity, let us consider only interactions expressed as a set of linear relationships with coefficients for each pairwise interaction  $C_{i,j}$ . (This simplification will be revisited below.)

Value created ( $V$ ) from time  $a$  to time  $b$  is then given by the integral:



$$V = \int_a^b \left( \sum_{i=1}^m f_i(I_i, \epsilon_i, t) + \sum_{j=1}^n e_j(\delta_j, t) + \sum_{i \neq j} C_{i,j} f_i(I_i, \epsilon_i, t) f_j(I_j, \epsilon_j, t) + \sum C_{i,j} f_i(I_i, \epsilon_i, t) e_j(\delta_j, t) + \sum_{i \neq j} C_{i,j} e_i(\delta_i, t) e_j(\delta_j, t) \right) dt$$

The percentage of total value created that is attributable to any actual investment  $i$  over any period of time is then 100 times the difference between this integral evaluated at the actual investment and this integral evaluated when that investment equals 0, all divided by the integral evaluated at the actual investment, or:

$$\% \text{ Credit for actual investment } i = \left( \frac{V_{|t=start, finish; investment_i=actual} - V_{|t=start, finish; investment_i=0}}{V_{|t=start, finish; investment_i=actual}} \right) * 100$$

The interaction terms can be made arbitrarily complicated, and this approach to percentage credit will still remain valid. Many of the interaction terms  $C_{i,j}$  would in reality be 0 since there would be many factors that in reality do not actually interact with each other.

This derivation shows that attributing to a particular infrastructure facility a percentage of the total value created by any endeavor that makes use of that facility is not just some loose, imaginary analogy. “Percentage credit” can be expressed as a precise mathematical formula, even if it is impossible in practice to fully express or evaluate this formula. It is thus sensible to talk about what might be reasonable values for the percentage of the credit for creating various kinds of value to attribute to any infrastructure facility supporting any sort of research or development. In the specific case of the XSEDE ecosystem, the function of investment in the XSEDE ecosystem would certainly interact with functions describing the impact of federal investment in LIGO and the COVID-19 pandemic. This approach is generally applicable to the idea of “percentage value attributable” for any facility, group, or individual contributing to any collaborative endeavor that creates value.

This derivation and a simple thought experiment show why it is nonsensical to take, as the value created by a piece of infrastructure, the sum of the values of all end products that are viewed as having depended on that piece of infrastructure for their creation. Imagine a data collection and analysis process that makes use of a sophisticated analytical instrument and an advanced computing facility. And suppose that the operators of the analytical instrument facility discuss a particular research end product with the primary investigator responsible for the creation of that end product. The researcher says that this research end product depended for its creation on this analytical instrument. Now suppose further that the operators of the advanced computing facility ask that same researcher whether this particular research

end product depended upon the advanced computing facility for its creation. The researcher, as can be easily imagined, again responds in the affirmative. Using this one research end product in tallying up the value attributable to the analytical instrument facility as well as in tallying up the value of the advanced computing facility means that the value of the research end product has just doubled. Its entire value was counted twice, once for each of the two research infrastructure facilities on which its creation depended.

**Acknowledgements** The authors would like to thank: Terry Campbell and Marlon Pierce for discussions of the ideas presented here; Dana Brunson for Campus Champions data; Scott Michael for help with the Appendix. Comments from anonymous reviewers and Robert Sinkovits improved this manuscript, as did editing by Marie Deer and Kristol Hancock. We thank everyone involved in the XSEDE ecosystem, particularly the many XSEDE staff and Campus Champions, who made XSEDE a success. We especially thank the researchers and students who used XSEDE’s services and gave it value. This material is based on work supported by the National Center for Atmospheric Research, a major facility sponsored by the NSF under Cooperative Agreement #1852977.

**Author Contributions** Conceptualization: all; Data curation: all; Formal analysis: Snapp-Childs, Hart, Costa, Stewart; Funding acquisition: Snapp-Childs, Towns; Investigation: all; Methodology: all; Administration, Resources, & Supervision: Snapp-Childs, Validation: all; Writing: all.

**Funding** University of Illinois Urbana Champaign National Center for Supercomputing Applications: NSF awards #1548562 & #2227627 (XSEDE & XSEDE2) & NSF award #2320345 (Delta). Pittsburgh Supercomputing Center: NSF awards #1445606 (Bridges) & #1928147 (Bridges2). Indiana University Pervasive Technology Institute: NSF awards #1445604 & #2005506. Purdue University Rosen Computing Center: NSF award #2005632 (Anvil). San Diego Supercomputer Center, University of California: NSF awards #0910847 (Gordon), #1341698 (Comet), & #1928224 (Expanse). University of Texas at Austin Texas Advanced Computing Center: NSF awards #1134872 (Stampede), #1341711 (Wrangler), & #1540931 (Stampede2). This material is based on work supported by the National Center for Atmospheric Research, a major facility sponsored by the NSF under Cooperative Agreement #1852977. Any opinions are those of the authors and may not represent the opinions of any funding agency including the NSF.

**Availability of data** Data are available from [5]. Data from XDMoD may be requested.

## Declarations

**Conflict of interest** A part of the salary of every author was paid for through NSF funding for the XSEDE ecosystem.

**Ethics approval** NA.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated

otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- National Academies of Sciences, Engineering, and Medicine: Information Technology Innovation: Resurgence, Confluence, and Continuing Impact. National Academies Press, Washington, DC (2020). <https://doi.org/10.17226/25961>.
- Wikipedia Community: Arctic Region Supercomputing Center. 2022. [https://en.wikipedia.org/wiki/Arctic\\_Region\\_Supercomputing\\_Center](https://en.wikipedia.org/wiki/Arctic_Region_Supercomputing_Center). Accessed 9 May 2024.
- Butrymowicz S, D'Amato P. Analysis: hundreds of colleges and universities show financial warning signs. The Hechinger Report. 2022. <https://hechingerreport.org/analysis-hundreds-of-colleges-and-universities-show-financial-warning-signs/>. Accessed 9 May 2024.
- Gardner L. Most public flagships are booming. Why are a handful flailing? The Chronicle of Higher Education. 2023. <https://www.chronicle.com/article/most-public-flagships-are-booming-why-are-a-handful-flailing?>. Accessed 9 May 2024.
- Costa CM, Wernert JA, Stewart CA, Snapp-Childs W, Blood PD, Chityala P, Hart D, Knepper R, McMullen DF, Mehringer S, Rogers G, Sinkovits R. Dataset: XSEDE return on investment data and analysis (2014–2022). Indiana University. 2024. <https://hdl.handle.net/2022/29584>. Accessed 9 May 2024.
- Towns J, Cockerill T, Dahan M, Foster I, Gaither K, Grimshaw A, Hazlewood V, Lathrop S, Lifka D, Peterson GD, Roskies R, Scott JR, Wilkins-Diehr N. XSEDE: accelerating scientific discovery. *Comput Sci Eng*. 2014;16:62–74. <https://doi.org/10.1109/mcse.2014.80>.
- XSEDE: XSEDE—The eXtreme Science and Engineering Discovery Environment (OAC 15-48562) Final Project Report September 1, 2016–August 31, 2023. <https://hdl.handle.net/2142/120470>. Accessed 9 May 2024.
- Stewart CA, Costa CM, Wernert JA, Snapp-Childs W, Bland M, Blood P, Campbell T, Couvares P, Fischer J, Hancock DY, Hart DL, Jankowski H, Knepper R, McMullen DF, Mehringer S, Pierce M, Rogers G, Sinkovits RS, Towns J. Use of accounting concepts to study research: return on investment in XSEDE, a US cyberinfrastructure service. *Scientometrics*. 2023;128(6):3225–55. <https://doi.org/10.1007/s11192-022-04539-8>.
- Stewart CA, Wernert JA, Costa CM, Hancock DY, Knepper R, Snapp-Childs W. Return on investment in research cyberinfrastructure: state of the art. In: Practice and experience in advanced research computing PEARC '22. Boston: ACM; 2022. p. 1–4. <https://doi.org/10.1145/3491418.3535131>.
- Apon AW, Ngo LB, Payne ME, Wilson PW. Assessing the effect of high performance computing capabilities on academic research output. *Empirical Economics*. 2015;48(1):283–312. <https://doi.org/10.1007/s00181-014-0833-7>.
- Smith P, Lien Harrell S. Research computing on campus-application of a production function to the value of academic high-performance computing. In: Practice and experience in advanced research computing. PEARC '21. New York: ACM; 2021. <https://doi.org/10.1145/3437359.3465564>.
- Smith P. Measuring the relative outputs of computational researchers in higher education. In: Practice and experience in advanced research computing. PEARC '22. New York: ACM; 2022. <https://doi.org/10.1145/3491418.3530764>.
- Kinney MR, Raiborn CA. Cost accounting: foundations and evolutions. Boston: Cengage Learning; 2012. <https://books.google.com/books?id=Nm3uBQP7cTIC>. Accessed 9 May 2024.
- Stewart CA, Hancock DY, Wernert J, Furlani T, Lifka D, Sill A, Berente N, McMullen DF, Cheatham T, Apon A, Payne R, Slavin SD. Assessment of financial returns on investments in cyberinfrastructure facilities: a survey of current methods. In: Practice and experience in advanced research computing. PEARC '19. 2019; p. 1–8. <https://doi.org/10.1145/3332186.3332228>.
- Stobierski T. How to calculate ROI to justify a project. Harvard Business School Online's Business Insights Blog. 2020. <https://online.hbs.edu/blog/post/how-to-calculate-roi-for-a-project>. Accessed 9 May 2024.
- WK Kellogg Foundation: WK Kellogg Foundation Logic Model Development Guide, Battle Creek, MI. 2004. [https://www.bette.revaluation.org/sites/default/files/2021-11/Kellogg\\_Foundation\\_Logic\\_Model\\_Guide.pdf](https://www.bette.revaluation.org/sites/default/files/2021-11/Kellogg_Foundation_Logic_Model_Guide.pdf). Accessed 9 May 2024.
- Del Bo CF. The rate of return to investment in R&D: the case of research infrastructures. *Technol Forecast Soc Chang*. 2016;112:26–37. <https://doi.org/10.1016/j.techfore.2016.02.018>.
- Stewart CA, Hancock DY, Wernert J, Link MR, Wilkins-Diehr N, Miller T, Gaither K, Snapp-Childs W. Return on investment for three cyberinfrastructure facilities: a local campus super-computer; the NSF-funded Jetstream cloud system; and XSEDE (the eXtreme Science and Engineering Discovery Environment). In: 2018 IEEE/ACM 11th International Conference on Utility and Cloud Computing (UCC). UCC'18. Washington, DC: IEEE Computer Society (2018). p. 223–236. <https://doi.org/10.1109/UCC.2018.00031>.
- The International Integrated Reporting Council: International IR Framework. 2021. <https://www.integratedreporting.org/resource/international-ir-framework/>. Accessed 9 May 2024.
- National Science Foundation: award abstract 1548562 XSEDE 2.0: integrating, enabling and enhancing national cyberinfrastructure with expanding community involvement. 2016. [https://www.nsf.gov/awardsearch/showAward?AWD\\_ID=1548562](https://www.nsf.gov/awardsearch/showAward?AWD_ID=1548562). Accessed 9 May 2024.
- Furlani TR, Schneider BL, Jones MD, Towns J, Hart DL, Gallo SM, DeLeon RL, Lu C-D, Ghadersohi A, Gentner RJ, Patra AK, von Laszewski G, Wang F, Palmer J.T, Simakov N. Using XDMoD to facilitate XSEDE operations, planning and analysis. In: Proceedings of the conference on eXtreme science and engineering discovery environment, XSEDE '13, San Diego, CA. 2013. p. 1–8. <https://doi.org/10.1145/2484762.2484763>.
- XSEDE: IDEALS @ Illinois: XSEDE: eXtreme science and engineering discovery environment. <https://www.ideals.illinois.edu/handle/2142/35660>. Accessed 9 May 2024.
- Chityala P, Costa C.M, Wernert JA, Stewart CA. Cyberinfrastructure value: a survey on perceived importance and usage. In: Practice and experience in advanced research computing PEARC '22. Boston: ACM; 2022. p. 1–4. <https://dl.acm.org/doi/10.1145/3491418.3535173>.
- ACM: ACM OPEN. 2023. <https://libraries.acm.org/subscriptions-access/acmopen>. Accessed 9 May 2024.
- Rousseau S, Catalano G, Daraio C. Can we estimate a monetary value of scientific publications? *Res Policy*. 2021;50(1):104116. <https://doi.org/10.1016/j.respol.2020.104116>.
- Mahendran S, Costa CM, Wernert JA, Stewart CA. Institutional value of a nobel prize. In: Practice and experience in advanced research computing (PEARC '22). Boston: ACM; 2022. p. 1–4. <https://doi.org/10.1145/3491418.3535157>.
- Nobel Foundation: Press Release. 2017. <https://www.nobelprize.org/prizes/physics/2017/press-release/>. Accessed 9 May 2024.

28. Swedish Krona to US Dollar Spot Exchange Rates for 2017. 2024. <https://www.exchangerates.org.uk/SEK-USD-spot-exchange-rates-history-2017.html>. Accessed 9 May 2024.
29. Krajec R. How much does a patent cost? BlueIron IP. 2023. <https://blueironip.com/how-much-does-a-patent-cost/>. Accessed 9 May 2024.
30. Boerner TJ, Deems S, Furlani TR, Knuth SL, Towns J. ACCESS: advancing innovation: NSF's advanced cyberinfrastructure coordination ecosystem: services & support. In: Practice and experience in advanced research computing. PEARC '23. New York: ACM; 2023. p. 173–176. <https://doi.org/10.1145/3569951.3597559>
31. XSEDE: URLs for XSEDE Annual or Project Year Reports. 2022. <https://hdl.handle.net/2142/114456>. Accessed 9 May 2024.
32. Snapp-Childs W, Costa C, Hart D, Olds D, Snell A, Wernert JA, Stewart CA. Value of practical training in high performance computing to stem hiring managers. To be submitted to Frontiers in Research Metrics and Analytics. 2024.
33. Day JC, Martinez A. Does majoring in STEM Lead to a STEM job after graduation? US Census Bureau. 2021. <https://www.census.gov/library/stories/2021/06/does-majoring-in-stem-lead-to-stem-job-after-graduation.html>. Accessed 9 May 2024.
34. Wilkins-Diehr N, Zentner M, Pierce M, Dahan M, Lawrence K, Hayden L, Mullinix N. The science gateways community Institute at two years. In: Proceedings of the practice and experience on advanced research computing. PEARC '18. New York: ACM. 2018. <https://doi.org/10.1145/3219104.3219142>.
35. Gesing S, Stirm C, Zentner M, Dahan M, Hayden L. SGX3: Novel concepts to enhance knowledge and extend the community around science gateways. In: Science gateways 2023 (SG23). Science Gateways (SG). 2023. p. 1–8. <https://doi.org/10.5281/zenodo.10034892>.
36. XSEDE: XSEDE—the eXtreme Science and Engineering Discovery Environment - Quarterly Report for 2014 Q4: October 1, 2014 through December 31, 2014. 2014. <http://hdl.handle.net/2142/78275>. Accessed 9 May 2024.
37. Miller MA, Schwartz T, Pfeiffer W. Embedding CIPRES science gateway capabilities in phylogenetics software environments. In: Proceedings of the conference on eXtreme science and engineering discovery environment. XSEDE '13. New York, ACM; 2013. <https://doi.org/10.1145/2484762.2484806>.
38. Zheng W, Zhang C, Bell EW, Zhang Y. I-TASSER gateway: a protein structure and function prediction server powered by XSEDE. Future Gener Comput Syst. 2019;99:73–85. <https://doi.org/10.1016/j.future.2019.04.011>.
39. Brase J, Campbell N, Helland B, Hoang T, Parashar M, Rosenfield M, Sexton J, Towns J. The COVID-19 high-performance computing consortium. Comput Sci Eng. 2022;24(1):78–85. <https://doi.org/10.1109/MCSE.2022.3145608>.
40. Brase J, Campbell N, Helland B, Hoang T, Parashar M, Rosenfield M, Sexton J, Towns J. The COVID-19 high-performance computing consortium. [https://s3.us-south.cloud-object-storage.appdomain.cloud/covid-19-hpc-object-storage-production/Consortium\\_Overview\\_Paper\\_03\\_2022\\_1f72939a70](https://s3.us-south.cloud-object-storage.appdomain.cloud/covid-19-hpc-object-storage-production/Consortium_Overview_Paper_03_2022_1f72939a70). Accessed 9 May 2024.
41. Keller E, Newman JE, Ortmann A, Jorm LR, Chambers GM. How much is a human life worth? A systematic review. Value Health. 2021;24(10):1531–41. <https://doi.org/10.1016/j.jval.2021.04.003>.
42. Hay M, Thomas DW, Craighead JL, Economides C, Rosenthal J. Clinical development success rates for investigational drugs. Nat Biotechnol. 2014;32:40–51. <https://doi.org/10.1038/nbt.2786>.
43. Pirtle B, Kimes R, McGovern A, Brown R. Using the XSEDE supercomputing and visualization resources to improve tornado prediction using data mining. In: Proceedings of the 1st conference of the eXtreme science and engineering discovery environment. XSEDE '12. New York: ACM; 2012. <https://doi.org/10.1145/2335755.2335831>.
44. National Weather Service: Weather Related Fatality and Injury Statistics, Washington, DC. 2022. <https://www.weather.gov/hazstat/>. Accessed 9 May 2024.
45. COVID-19 HPC Consortium: Home page. 2022. <https://covid19-hpc-consortium.org/>. Accessed 9 May 2024.
46. Towns J, Hart DL. Bibliography of publications enabled by use of the COVID-19 HPC consortium. Published by the National Center for Supercomputing Applications, University of Illinois Urbana-Champaign. 2024. <https://hdl.handle.net/2142/122297>. Accessed 9 May 2024.
47. Machi D, Bhattacharya P, Hoops S, Chen J, Mortveit H, Venkatraman S, Lewis B, Wilson M, Fadikar A, Maiden T, Barrett CL, Marathe MV. Scalable epidemiological workflows to support COVID-19 planning and response. In: 2021 IEEE International Parallel and Distributed Processing Symposium (IPDPS). Los Alamitos: IEEE Computer Society, 2021. p. 639–650. <https://doi.org/10.1109/IPDPS49936.2021.00072>
48. Howerton E, Contamin L, Mullany LC, Qin M, Reich NG, Bents S, Borchering RK, Jung S-M, Loo SL, Smith CP, Levander J, Kerr J, Espino J, Panhuis WG, Hochheiser H, Galanti M, Yamana T, Pei S, Shaman J, Rainwater-Lovett K, Kinsey M, Tallaksen K, Wilson S, Shin L, Lemaître JC, Kaminsky J, Hulse JD, Lee EC, McKee CD, Hill A, Karlen D, Chinazzi M, Davis JT, Mu K, Xiong X, Piontti A, Vespignani A, Rosenstrom ET, Ivy JS, Mayorga ME, Swann JL, España G, Cavany S, Moore S, Perkins A, Hladish T, Pillai A, Ben Toh K, Longini I, Chen S, Paul R, Janies D, Thill J-C, Bouchnita A, Bi K, Lachmann M, Fox SJ, Meyers LA, Srivastava A, Porebski P, Venkatraman S, Adiga A, Lewis B, Klahn B, Outten J, Hurt B, Chen J, Mortveit H, Wilson A, Marathe M, Hoops S, Bhattacharya P, Machi D, Cadwell BL, Healy JM, Slayton RB, Johansson MA, Biggerstaff M, Truelove S, Runge MC, Shea K, Viboud C, Lessler J. Evaluation of the US COVID-19 scenario modeling hub for informing pandemic response under uncertainty. Nat Commun. 2023;14(7260):1–15. <https://doi.org/10.1038/s41467-023-42680-x>.
49. Bhattacharya P, Chen J, Hoops S, Machi D, Lewis B, Venkatraman S, Wilson ML, Klahn B, Adiga A, Hurt B, Outten J, Adiga A, Warren A, Baek YY, Porebski P, Marathe A, Xie D, Swarup S, Vullikanti A, Mortveit H, Eubank S, Barrett CL, Marathe M. Data-driven scalable pipeline using national agent-based models for real-time pandemic response and decision support. Int J High Perform Comput Appl. 2023;37(1):4–27. <https://doi.org/10.1177/10943420221127034>.
50. IIT Madras: Prof. Madhav Marathe: “Never give up; life is much more than GPAs and ranks.” Indian Institute of Technology, Madras (2023). <https://joyofgiving.alumni.iitm.ac.in/alma-matres/alumni-spotlight/prof-madhav-marathe-never-give-up-life-is-much-more-than-gpas-and-ranks>. Accessed 9 May 2024.
51. National Center for Health Statistics: COVID-19 Mortality by State. National Center for Health Statistics, Centers for Disease Control and Prevention. 2023. [https://www.cdc.gov/nchs/pressroom/sosmap/covid19\\_mortality\\_final/COVID19.htm](https://www.cdc.gov/nchs/pressroom/sosmap/covid19_mortality_final/COVID19.htm). Accessed 9 May 2024.
52. Ahmad FB, Cisewski JA. Quarterly provisional estimates for selected indicators of mortality, 2018–Quarter 3, 2020. National Center for Health Statistics. National Vital Statistics System, Vital Statistics Rapid Release Program. 2021. [https://www.cdc.gov/nchs/pressroom/sosmap/covid19\\_mortality/COVID\\_12\\_Months.htm](https://www.cdc.gov/nchs/pressroom/sosmap/covid19_mortality/COVID_12_Months.htm). Accessed 9 May 2024.
53. Centers for Disease Control and Prevention: Deaths by Week and State—Provisional Death Counts for COVID-19. 2023. <https://www.cdc.gov/nchs/nvss/vsrr/covid19/index.htm>. Accessed 9 May 2024.



54. Yakusheva O, van den Broek-Altenburg E, Brekke G, Atherly A. Lives saved and lost in the first six month of the US COVID-19 pandemic: a retrospective cost-benefit analysis. *PLoS One*. 2022. <https://doi.org/10.1371/journal.pone.0261759>.
55. Chen J, Vullikanti A, Santos J, Venkatramanan S, Hoops S, Mortveit H, Lewis B, You W, Eubank S, Marathe M, Barrett C, Marathe A. Epidemiological and economic impact of COVID-19 in the US. *Sci Rep*. 2021;11:2045. <https://doi.org/10.1038/s41598-021-99712-z>.
56. Cutler DM, Summers LH. The COVID-19 pandemic and the \$16 Trillion Virus. *JAMA*. 2020;324:1495–6. <https://doi.org/10.1001/jama.2020.19759>.
57. National Science Foundation: SCC-IRG Track 1: Preparing for Future Pandemics: Subway Crowd Management to Minimize Airborne Transmission of Respiratory Viruses (Way-CARE). Award Abstract # 2218809. 2023. [https://www.nsf.gov/awardsearch/showAward?AWD\\_ID=2218809](https://www.nsf.gov/awardsearch/showAward?AWD_ID=2218809). Accessed 9 May 2024.
58. Florio M, Sirtori E. The evaluation of research infrastructures: a cost-benefit analysis framework. Milan European Economic Workshop. 2014. <https://doi.org/10.2139/ssrn.2722500>.
59. Castelvechi D. Black hole at the centre of our Galaxy imaged for the first time. *Nature*. 2022;605(7910):403–4. <https://doi.org/10.1038/d41586-022-01320-y>.
60. Valero MM, Jofre L, Torres R. Multifidelity prediction in wildfire spread simulation: modeling, uncertainty quantification and sensitivity analysis. *Environ Model Softw*. 2021;141: 105050. <https://doi.org/10.1016/j.envsoft.2021.105050>.
61. Thomas D, Butry D, Gilbert S, Webb D, Fung J. The costs and losses of wildfires. Special Publication (NIST SP), National Institute of Standards and Technology, Gaithersburg; 2017. <https://doi.org/10.6028/NIST.SP.1215>
62. Callahan CW, Mankin JS. Globally unequal effect of extreme heat on economic growth. *Sci Adv*. 2022;8(43):3726. <https://doi.org/10.1126/sciadv.add3726>.
63. Marvin DC, Cameron D, Nelson E, Plantinga A, Breck J, Sencan G, Passero M. Toward a carbon neutral california: economic and climate benefits of land use interventions. Next 10, San Francisco; 2018. [https://next10.org/sites/default/files/toward-carbon-neutral-california\\_1.pdf](https://next10.org/sites/default/files/toward-carbon-neutral-california_1.pdf). Accessed 9 May 2024.
64. O'Hara M, Yao C, Ye M. What's not there: odd lots and market data. *J Finance*. 2014;69(5):2199–236. <https://doi.org/10.1111/jofi.12185>.
65. Goldstein I, Spatt CS, Ye M. Big data in finance. *Rev Financ Stud*. 2021;34(7):3213–25. <https://doi.org/10.1093/rfs/hhab038>.
66. Ye M. Big data in finance. National Bureau of Economic Research Lecture: Summer Institute Big Data and High-Performance Computing for Financial Economics. 2018. <https://www.youtube.com/watch?v=fw1WOMlhCuE>. Accessed 9 May 2024.
67. Mukunda G. The social and political costs of the financial crisis, 10 years later. *Harvard Business Review*. 2018. <https://hbr.org/2018/09/the-social-and-political-costs-of-the-financial-crisis-10-years-later>. Accessed 9 May 2024.
68. Subcommittee on Networking and Information Technology Research and Development, Committee on Science and Technology Enterprise and the Subcommittee on Future Advanced Computing Ecosystem, Committee on Technology of the National Science and Technology Council: National Strategic Computing Reserve: a Blueprint. Technical report. 2021. <https://www.whitehouse.gov/wp-content/uploads/2021/10/National-Strategic-Computing-Reserve-Blueprint-Oct2021.pdf>. Accessed 9 May 2024.
69. Florio M, Forte S, Sirtori E. Forecasting the socio-economic impact of the Large Hadron Collider: a cost-benefit analysis to 2025 and beyond. *Technol Forecast Soc Change*. 2016;112:38–53. <https://doi.org/10.1016/j.techfore.2016.03.007>.
70. Wikipedia Community: Shadow Price. 2023. [https://en.wikipedia.org/wiki/Shadow\\_price](https://en.wikipedia.org/wiki/Shadow_price). Accessed 9 May 2024.
71. Carpenter SR, Turner M. Opening the black boxes: ecosystem science and economic valuation. *Ecosystems*. 2000;3(1):1–3. <https://doi.org/10.1007/s100210000001>.
72. Starrett DA. Shadow pricing in economics. *Ecosystems*. 2000;3:16–20. <https://doi.org/10.1007/s100210000004>.
73. Frontier Economics: Rates of return to investment in science and innovation. 2014. <https://www.gov.uk/government/publications/science-and-innovation-rates-of-return-to-investment>. Accessed 9 May 2024.
74. Bureau of Economic Analysis: RIMS II: An essential tool for regional developers and planners. 2013. [https://www.bea.gov/sites/default/files/methodologies/RIMSII\\_User\\_Guide.pdf](https://www.bea.gov/sites/default/files/methodologies/RIMSII_User_Guide.pdf). Accessed 9 May 2024.
75. IMPLAN Group LLC.: IMPLAN. Home page. 2024. <https://implan.com/>. Accessed 9 May 2024.
76. Beagrie N. The Value and Impact of the European Bioinformatics Institute. Charles Beagrie, Ltd. Salisbury; 2016. <http://www.beagrie.com/EBI-impact-report.pdf>. Accessed 9 May 2024.
77. Sullivan KP, Brennan-Tonetta P, Marxen LJ. Economic impacts of the research collaborative for structural bioinformatics (RCSB) protein data bank. Rutgers University; 2017. [https://doi.org/10.2210/rcsb\\_pdb/pdb-econ-imp-2017](https://doi.org/10.2210/rcsb_pdb/pdb-econ-imp-2017)
78. Ribes D, Lee CP. Sociotechnical studies of cyberinfrastructure and e-research: current themes and future trajectories. *Comput Supp Coop Work (CSCW)*. 2010;19:231–44. <https://doi.org/10.1007/s10606-010-9120-0>.
79. Stewart CA, Roskies R, Knepper R, Moore R.L, Whitt J, Cockerill TM. XSEDE value added, cost avoidance, and return on investment. In: Proceedings of the 2015 XSEDE conference. XSEDE '15 conference, St. Louis, MO; 2015. p. 1–8. <https://doi.org/10.1145/2792745.2792768>
80. Newby G, Apon AW, Berente N, Eigenmann R, Fratkin S, Lifka DA, Stewart CA. Return on investment from academic supercomputing. Panel session at SC14—the international conference for high performance computing, networking, storage and analysis. New Orleans, LA; 2014. <http://hdl.handle.net/2022/19242>. Accessed 9 May 2024.
81. Lane J. Assessing the impact of science funding. *Science*. 2009;324:1273–5. <https://doi.org/10.1126/science.1175335>.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.