Adventures in Communicating the Value of Campus CI Investment

Preston Smith

Throughput Computing 2023 – July 11, 2023

PURDUE UNIVERSITY®

Plug: This is the abridged version!

Catch me in Chicago this fall for the full version!

Campus Cyberinfrastucture

A Coordinated Investment by Many Stakeholders

NSF and other agencies

Campuses

- CIOs
- VPRs
- Faculty

My talk will focus mostly on the latter!

PURDUE UNIVERSITY

Communicating Value

Hot Take: Our Community is Immature at Communicating Value

Questions heard at Purdue from finance leaders:

"What do they keep spending all that money on for supercomputers?"

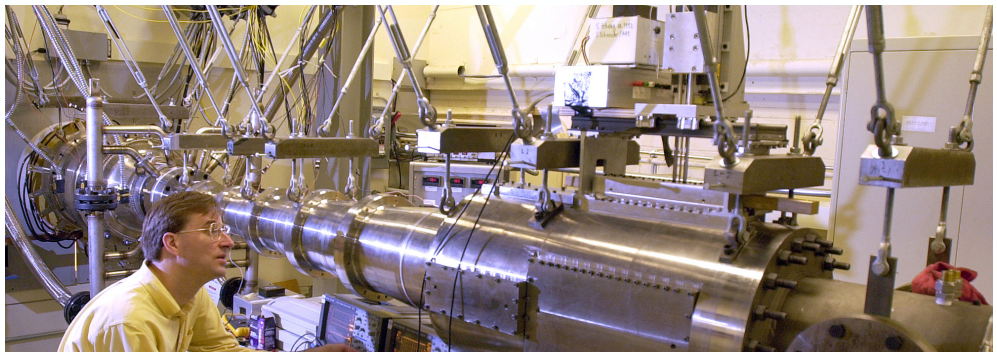
"Who even uses these things? Are they just vanity toys for a couple of people?"

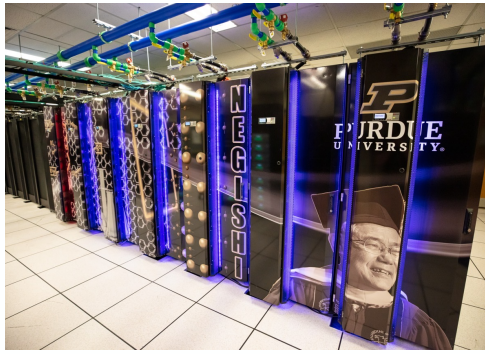
"Why is that recharge center not covering its costs?"

PURDUE UNIVERSITY

Motivation

Value of HPC investments





Computing systems are critical pieces of research infrastructure

Administrators have to manage competing priorities of infrastructures in which to invest:

- Supercomputers
- Wind tunnels
- Electron Microscopes
- Enterprise IT vs research IT
- Etc

How to demonstrate that computational research is a good investment?

PURDUE UNIVERSITY

Step 1: Know your Stakeholders

Which specific stakeholders are you communicating value to?

What is important at the institution at this point in time?

Faculty: getting next grant, publishing, graduating students

Deans, Heads, Chairs: Students, educational outcomes, Faculty recruitment and retention, Having appropriate infrastructure for their units to succeed

Provost: Students, educational outcome, happy faulty

VP for Research: Competitiveness for funding, winning major grants and contracts, growing R&D expenditures, tech transfer, IP

CFO: $$$ - ROI, efficiency, reduction of expense, increase of revenue

CIO: Contributing to the mission

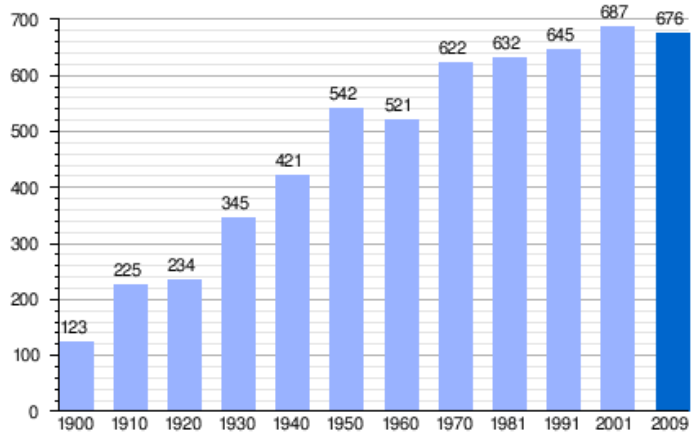
Personas

Stakeholder Interests

Step 2: Have Quantitative Center Metrics

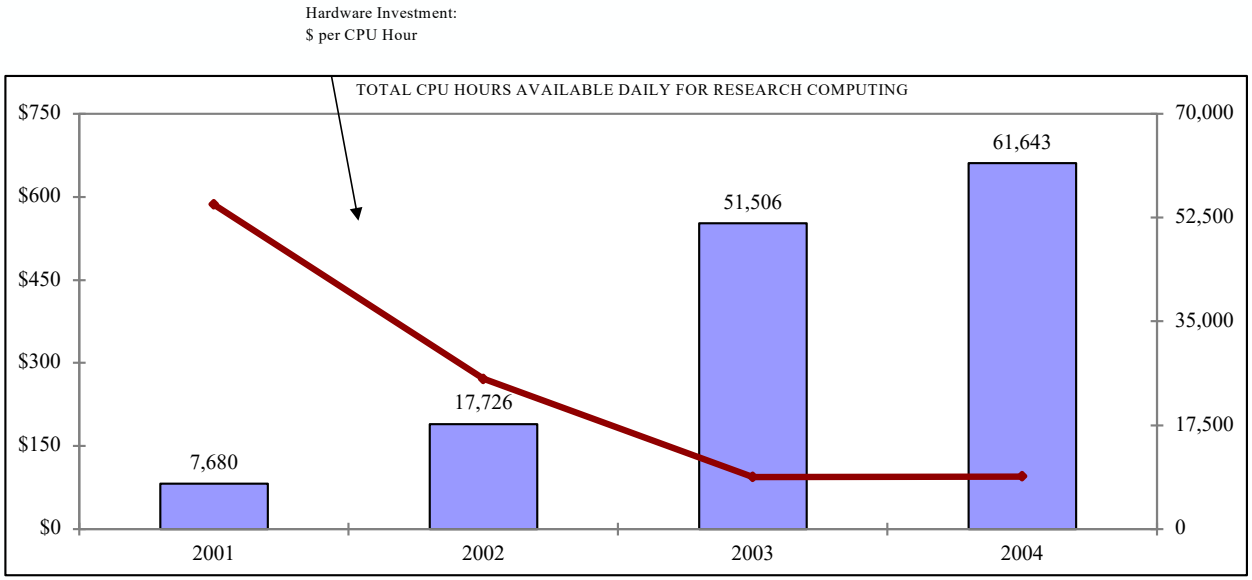
Telling your story

.. based on the interests of the specific stakeholder, given the institution's priorities in the current context



To All Stakeholders:

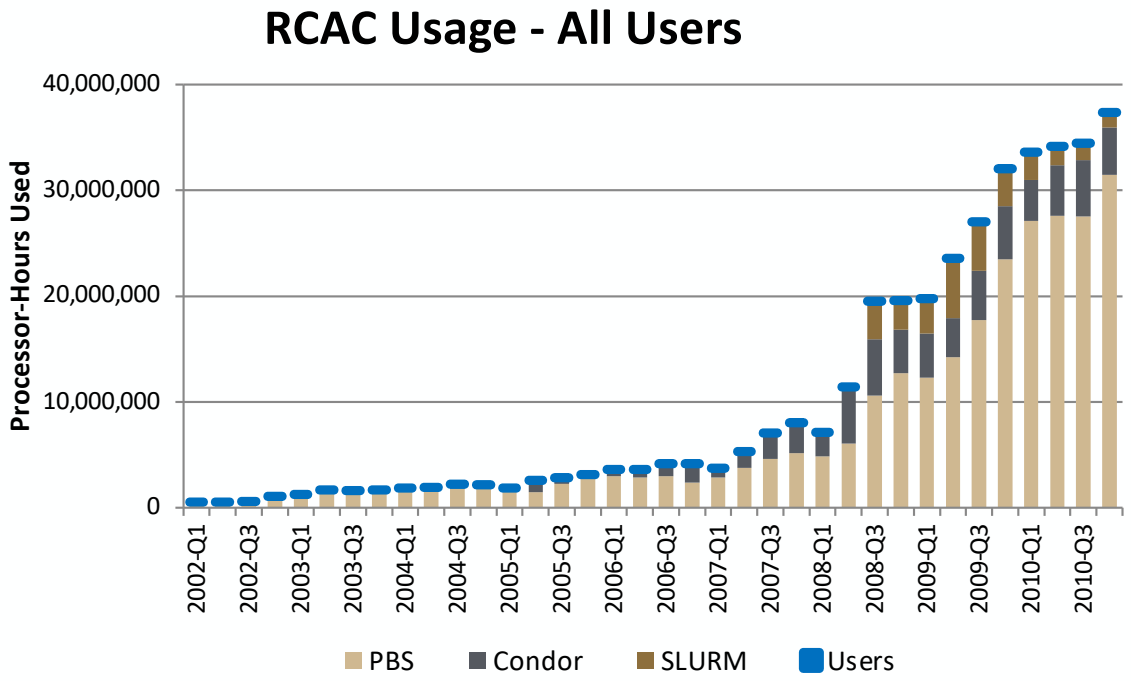
We are addressing the ~2001 lack of capacity and driving costs down



PURDUE UNIVERSITY

To All Stakeholders:

The capacity that we've invested in is well-used



124% average annual growth on the number of computational hours used by Purdue faculty

PURDUE UNIVERSITY

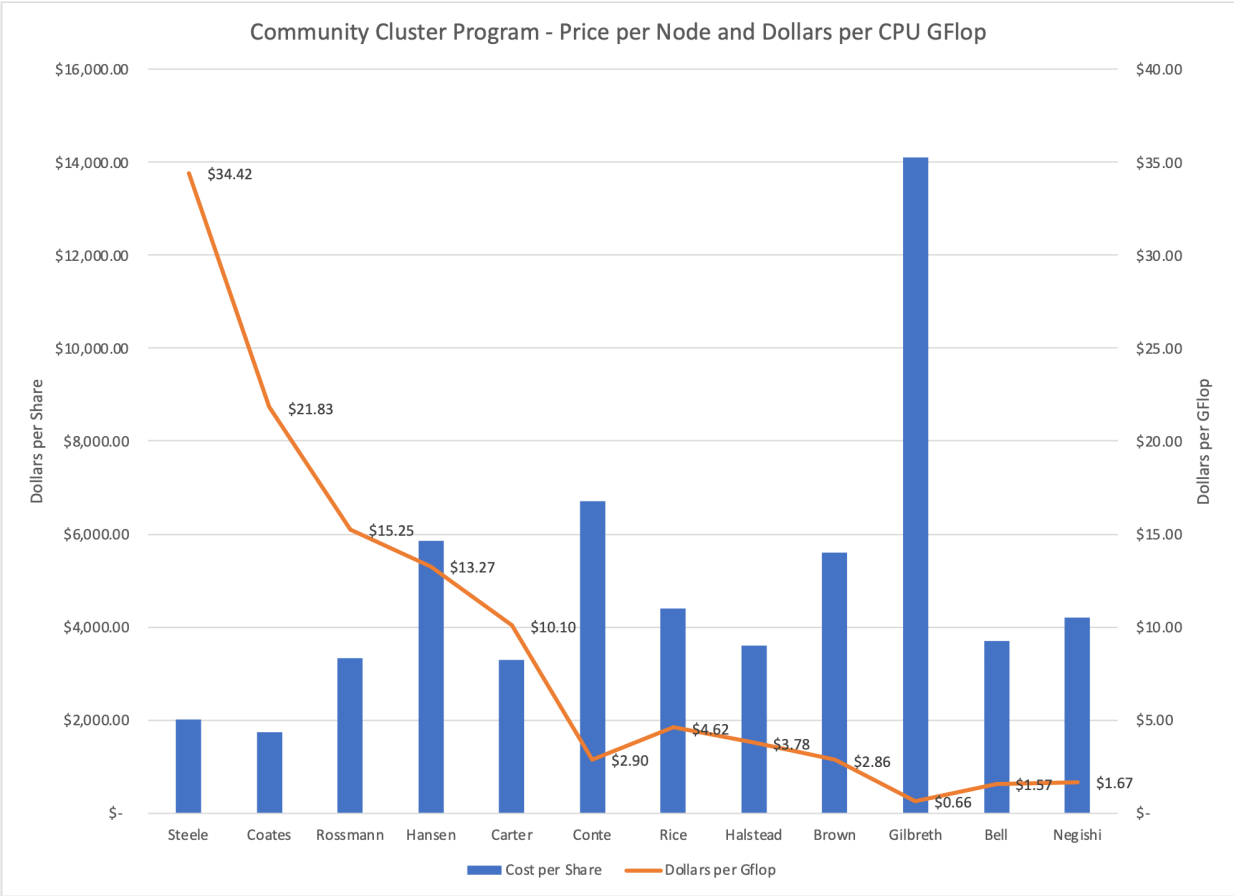
Grid computing: 30M↑
from 18M hours (2008)

Storage: 2,588 TB↑
from 1,904 TB of capacity (2009)

HPC: 113M↑
from 67M hours (2009)

To Faculty and CFO:

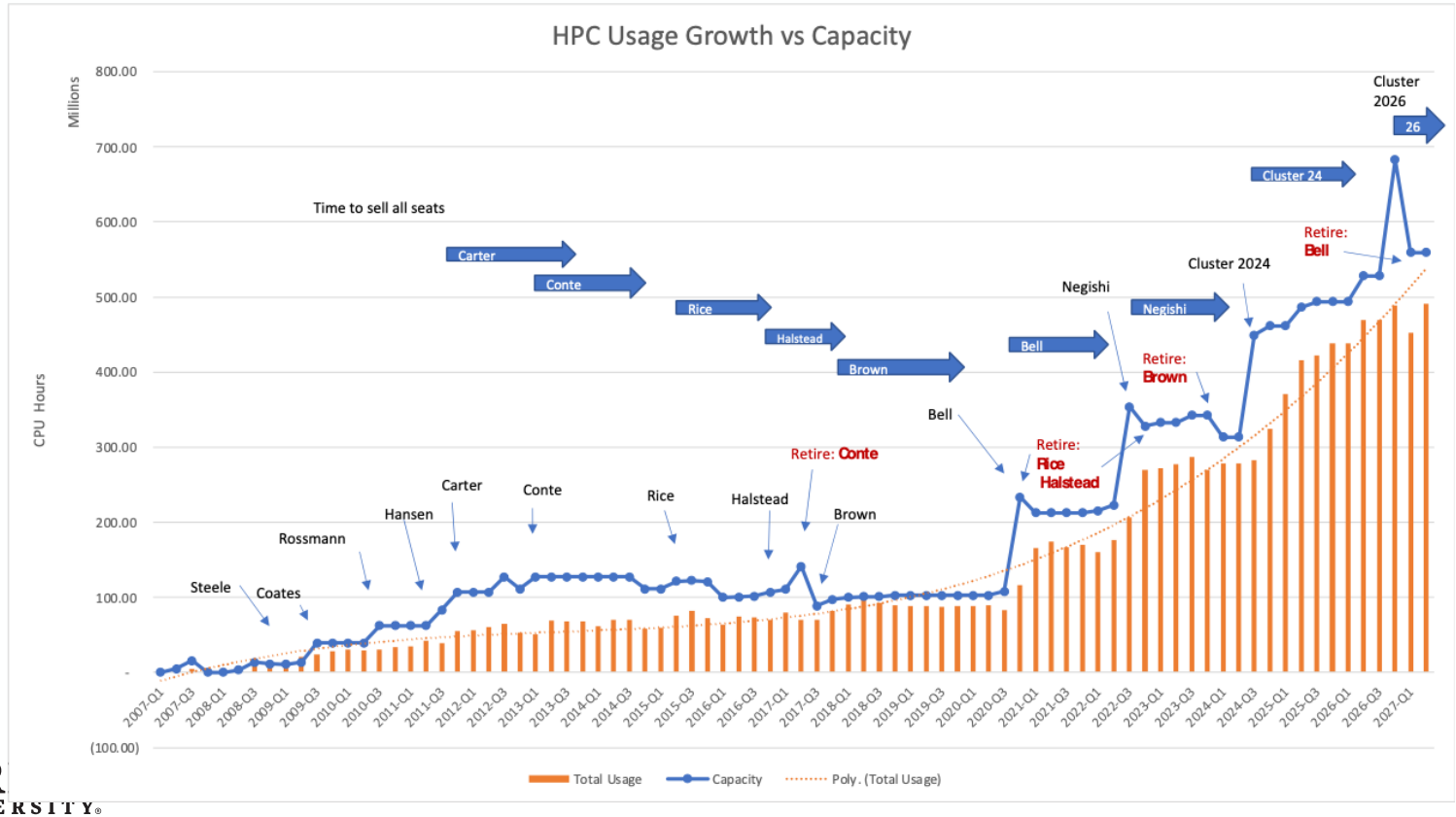
Each iteration gets us more capability per dollar



PURDUE UNIVERSITY

To CIO and CFO:

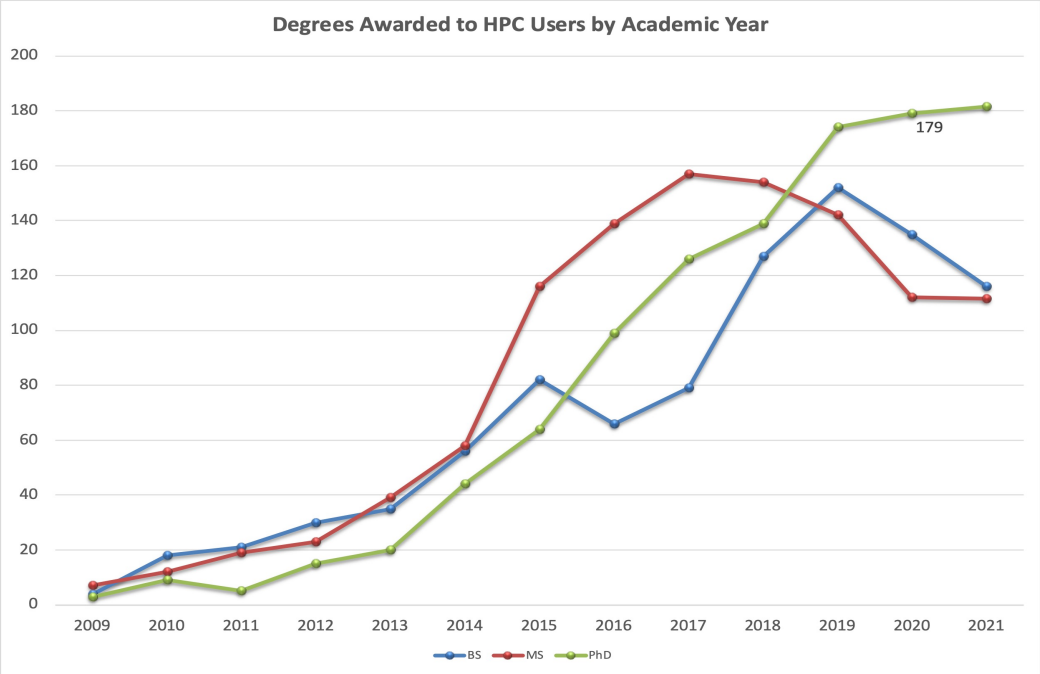
We understand our sales and usage patterns well enough to plan our lifecycle needs



PURDUE UNIVERSITY

To Faculty and Academic Leaders

To train students, CI resources are critical tools to have available



Year	Earned Doctorates	HPC-Using Doctorates	% Using HPC
2010	639	9	1%
2011	672	5	1%
2012	656	15	2%
2013	687	20	3%
2014	735	44	6%
2015	709	64	9%
2016	727	99	14%
2017	740	127	17%
2018	758	140	18%
2019	738	175	24%
2020	808	182	23%
2021	802	204	25%
2022	835	205	25%

PURDUE UNIVERSITY

To the CFO:

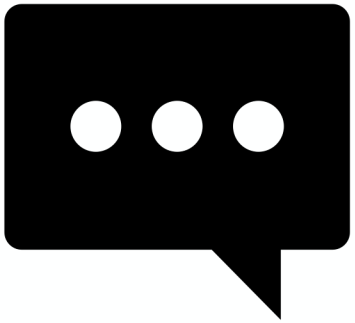
The general fund costs for campus CI enable research expenditures have a high ROI

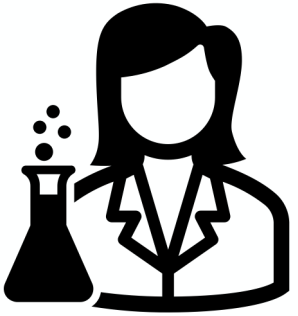
2022 Metric	Amount ($M)
Net Cost to Purdue	$5.21
Awards to RCAC-using Faculty	$350.80
Direct Expenditures Enabled	$329.10
F&A Expenditures Enabled	$69.90
Return on Investment	
Awards Return	67.33
Direct Expenditure Return	63.17
F&A Expenditure Return	13.42

PURDUE UNIVERSITY

Step 3: Get Testimonials

Don’t underestimate the power of qualitative data





Before: (~2006)

"I'd rather remove my appendix with a spork than let you people run my research computers."

PURDUE UNIVERSITY

Recruitment:

"Knowing there was a good group of experienced professionals I could rely on for support and establishing the computational infrastructure that I needed was very comforting when I was considering coming to Purdue. It frees up my time and the time of my graduate students and post-docs. We can focus on the scientific problems, which are our primary interest."

— Jeffrey Greeley

Charles and Nancy Davidson
Professor of Chemical Engineering

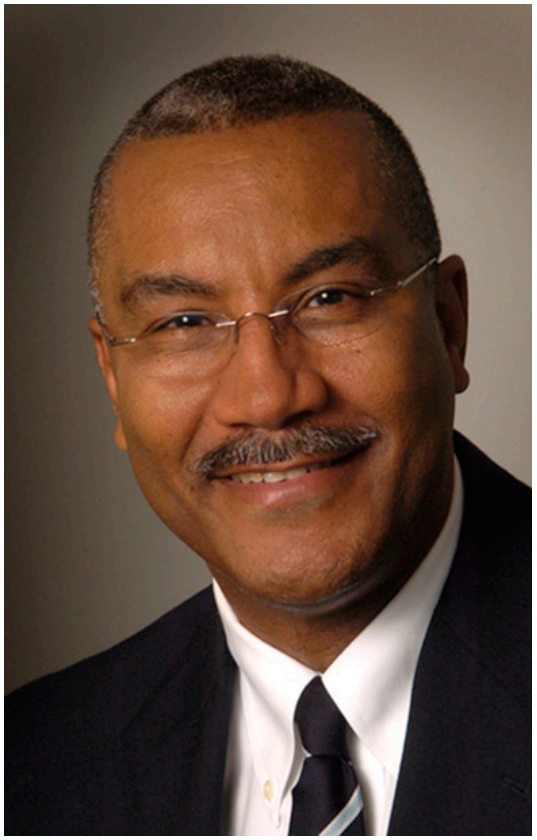
PURDUE UNIVERSITY

Enable Top Faculty

"I wouldn't have been elected to the National Academy of Sciences without these clusters. Having the clusters, we were able to set a very high standard that led a lot of people around the world to use our work as a benchmark, which is the kind of thing that gets the attention of the national academy."

— Joseph Francisco

William E. Moore Professor of Physical Chemistry, member of the National Academy of Sciences and past president of the American Chemical Society



PURDUE UNIVERSITY

Enable the Impossible

“We've been running things on the Conte cluster that would have taken months to run in a day. It's been a huge enabling technology for us.”

— Charles Bouman

Showalter Professor of Electrical and Computer Engineering and Biomedical Engineering

"It’s great to have world-class HPC systems like Bell at Purdue. Without these systems, we would not have been able to finish this study estimating the total number of tree species worldwide.”

— Jingjing Liang

Assistant Professor of Quantitative Forest Ecology

PURDUE UNIVERSITY

Production Function Model

Some guy did some research on the value proposition…

THE VALUE PROPOSITION OF CAMPUS HIGH PERFORMANCE COMPUTING FACILITIES TO INSTITUTIONAL PRODUCTIVITY - A PRODUCTION FUNCTION MODEL

by

Preston Smith

A Dissertation

Submitted to the Faculty of Purdue University

In Partial Fulfillment of the Requirements for the degree of

Doctor of Philosophy

Purdue Polytechnic Institute

West Lafayette, Indiana

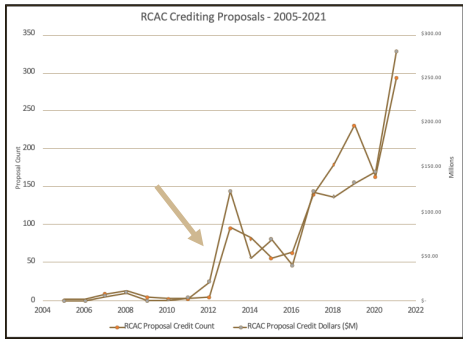
August 2022

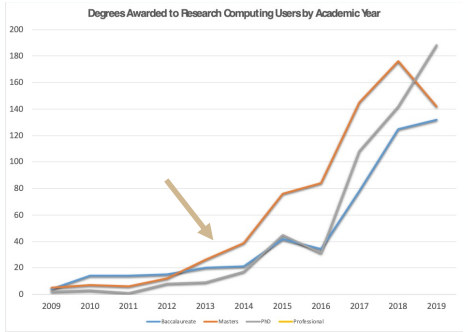
Motivator: Metrics at Purdue

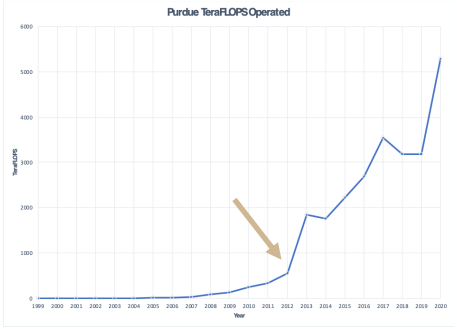
- CI-using awards, degrees, proposals all take huge jumps in 2012-2014
- ***What happened there?***
- Carter (our first big cluster) came online in late 2011.
- Conte (#28 on Top 500) in early 2013

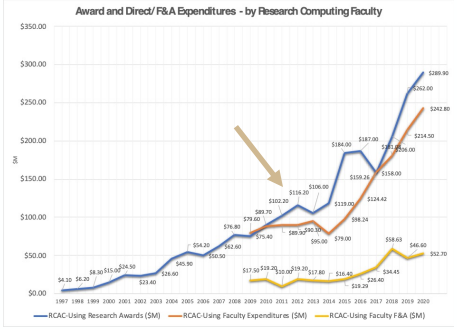
This prompted questions to me – did these investments contribute to increase in outputs?

Understanding these relationships is critical to modeling the impact of the investment!









PURDUE UNIVERSITY

Research Questions

- Does institutional investment in campus CI facilities lead to a measurable impact on institutional output?

- If so, what are the key factors for investment that lead to returns on institutional outputs?

- What is a model for quantifying the impact of campus cyberinfrastructure investments?

PURDUE UNIVERSITY

Background

Work by Amy Apon

(2010): Presence of a Top 500 supercomputer on a campus (an input) positively impacts research output at the institutional level.

(2015): Departments in Chemistry, Civil and Environmental Engineering, and Physics are more productive in universities with local cyberinfrastructure.

- **Downsides:** Top 500 is no longer a viable proxy for investment.
- How to add a labor dimension in model? No cross-institutional data available as to investment.

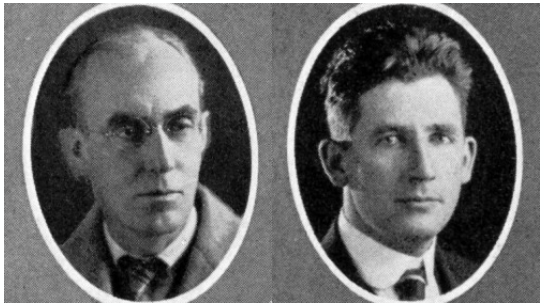
PURDUE UNIVERSITY

Production Function Model

In the 1920s, theories of production described the amount produced depended upon the level of technological knowledge and the quantities of the factors of production.

Cobb and Douglas (1927) described a least squares regression equation that predicts production (Y) based on inputs of capital (K) and labor (L),

$$Y = f(L^k K^{1-k})$$



"We found the values of k and 1-k by the method of least squares to be .75 and .25…"
Paul Douglas, Autobiography (1971)

PURDUE UNIVERSITY

Inputs and Outputs

What to use for Labor, Capital, and Outputs?

Table 3: Selected Purdue Input and Output Metrics

Input	Output
Annual net cost of the center	Purdue direct sponsor expenditures
Annual salary costs	Purdue F&A expenditures
Total TF operated	HERD reported R&D Expenditures
# of cluster faculty PIs	Purdue sponsored research awards
Center capital costs	Earned doctorates reported to NSF
	Purdue Publications
	High Impact Pubs
	US News ranking

PURDUE UNIVERSITY

Correlation Analysis

For Single Inputs TF and Salary

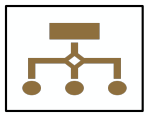
Output	Kendall Tau (TF)	Kendall Tau (Salary)
HERD Exp	0.92	0.88
Awards	0.83	0.75
Earned Doctorates	0.90	0.77
Purdue Pubs	0.90	0.32
Hi Impact Pubs	0.85	0.87

PURDUE UNIVERSITY

Regression Models

With Multiple Inputs Predicting an Output

$$Y = f(K, L) = K_{\text{flops}} L_{\text{staff}}$$



With the proxies for labor and capital identified, use the production function form **(labor + capital)** to model the various outputs.

For example, TeraFLOPS + staff salaries to predict outputs

PURDUE UNIVERSITY

Output – HERD Expenditures

Input	HERD Exp ($M)	% Variance Explained
100 TeraFLOPS	2.59	25%
$100k Salaries	9.04	43%
$100k RCAC Grants	2.34	28%

PURDUE UNIVERSITY

Output – Earned Doctorates

Input	Earned Doctorates	% Variance Explained
100 TeraFLOPS	2.55	31%
$100k Salaries	7.36	42%
$100k RCAC Grants	1.27	22%

PURDUE UNIVERSITY

Implications for Policymakers

Results, Summarized

Investments in both systems and people lead to measurable returns to the institution

Labor accounts for the largest amount of the variation in all models

When optimizing for publication-based outputs, salary investment yields strong returns

Allowing (or encouraging) HPC center staff to be PIs on their own grants or with faculty yields dividends to the institution

PURDUE UNIVERSITY

Broader Applications

How can this framework be applied to other CIs?

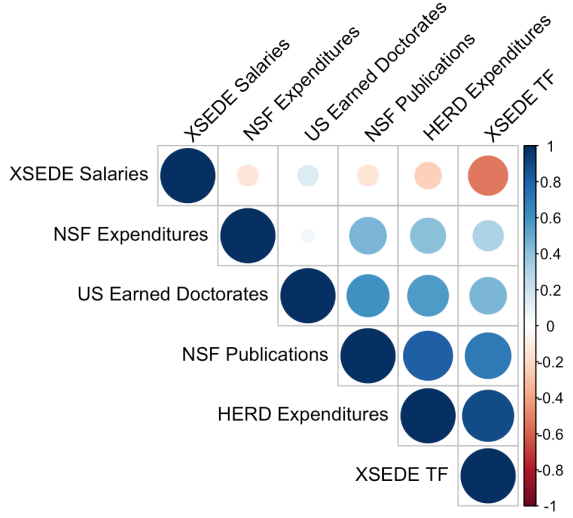
- Work ongoing to apply model to other campus' data.
  - *Can I talk to you about your data?*
- Work ongoing to apply model to NSF XSEDE infrastructure
  - Using all NSF expenditures, US Earned Doctorates, NSF-funded pubs in Scopus, etc.
  - Gap: only have XSEDE2 data, and lack staff costs that are in resource awards

Q: How could we adapt this model for OSG?

- Proxies for capacity?
- How to measure labor?
- Is there a quantifiable value to making your resources a part of the shared national CI?

term <chr>	XSEDE TF <dbl>
HERD Expenditures	0.7142857
NSF Expenditures	0.6190476
NSF Publications	0.4285714
US Earned Doctorates	-0.2380952
XSEDE Salaries	-0.5238095

5 rows



PURDUE UNIVERSITY

Thank You

psmith@purdue.edu

PURDUE UNIVERSITY®