

# *Adventures in Communicating the Value of Campus CI Investment*

Preston Smith

Metrics 2023

Denver, CO

November 11, 2023



## *Problem: finding investment for your Campus CI?*

The federal government has several programs to consider.. NSF CC\* is excellent for campus resources.

Also MRI, instrumentation programs from NIH (S10), DOD (DURIP), etc.

*“The proposal is expected to document long term commitment to operations and maintenance (O&M) past the lifetime of the award - or describe the strategy for sustaining availability of the resource.”*

COVER SHEET FOR PROPOSAL TO THE NATIONAL SCIENCE FOUNDATION											
PROGRAM ANNOUNCEMENT/SOLICITATION NO./DUE DATE <b>NSF 16-509</b>		<input checked="" type="checkbox"/> Special Exception to Deadline Date Policy		FOR NSF USE ONLY <b>NSF PROPOSAL NUMBER</b>							
FOR CONSIDERATION BY NSF ORGANIZATION UNIT(S) (Indicate the most specific unit known, i.e., program, division, etc.) <b>DEB - Long-Term Ecological Research</b>											
DATE RECEIVED	NUMBER OF COPIES	DIVISION ASSIGNED	FUND CODE	DUNS# (Data Universal Numbering System) <b>0748118034567</b>	FILE LOCATION						
EMPLOYER IDENTIFICATION NUMBER (EIN) OR TAXPAYER IDENTIFICATION NUMBER (TIN) <b>530206152</b>		SHOW PREVIOUS AWARD NO. IF THIS IS <input type="checkbox"/> A RENEWAL <input type="checkbox"/> AN ACCOMPLISHMENT-BASED RENEWAL		IS THIS PROPOSAL BEING SUBMITTED TO ANOTHER FEDERAL AGENCY? YES <input type="checkbox"/> NO <input checked="" type="checkbox"/> IF YES, LIST ACRONYM(S)							
NAME OF ORGANIZATION TO WHICH AWARD SHOULD BE MADE <b>National Science Foundation</b>		ADDRESS OF Awardee ORGANIZATION, INCLUDING 9 DIGIT ZIP CODE <b>National Science Foundation 4201 Wilson Boulevard Arlington, VA. 222301000</b>									
AWARDEE ORGANIZATION CODE (IF KNOWN) <b>4102852000</b>											
NAME OF PRIMARY PLACE OF PERF <b>ProdValid</b>		ADDRESS OF PRIMARY PLACE OF PERF, INCLUDING 9 DIGIT ZIP CODE <b>ProdValid AA.</b>									
IS Awardee ORGANIZATION (Check All That Apply) (See GPG II.C For Definitions)		<input type="checkbox"/> SMALL BUSINESS <input type="checkbox"/> FOR-PROFIT ORGANIZATION		<input type="checkbox"/> MINORITY BUSINESS <input type="checkbox"/> WOMAN-OWNED BUSINESS							
		<input type="checkbox"/> IF THIS IS A PRELIMINARY PROPOSAL THEN CHECK HERE									
TITLE OF PROPOSED PROJECT <b>SE ProdValid Jenkins Test</b>											
REQUESTED AMOUNT \$ <b>4,444</b>		PROPOSED DURATION (1-60 MONTHS) <b>24 months</b>		REQUESTED STARTING DATE <b>12/12/16</b>							
		SHOW RELATED PRELIMINARY PROPOSAL NO. IF APPLICABLE									
THIS PROPOSAL INCLUDES ANY OF THE ITEMS LISTED BELOW <input type="checkbox"/> BEGINNING INVESTIGATOR (GPG I.G.2) <input type="checkbox"/> DISCLOSURE OF LOBBYING ACTIVITIES (GPG II.C.1.e) <input type="checkbox"/> PROPRIETARY & PRIVILEGED INFORMATION (GPG I.D, II.C.1.d) <input type="checkbox"/> HISTORIC PLACES (GPG II.C.2.j) <input type="checkbox"/> VERTEBRATE ANIMALS (GPG II.D.6) IACUC App. Date _____ PHS Animal Welfare Assurance Number _____ <input checked="" type="checkbox"/> FUNDING MECHANISM <b>Research - other than RAPID or EAGER</b>						<input type="checkbox"/> HUMAN SUBJECTS (GPG II.D.7) Human Subjects Assurance Number _____ Exemption Subsection _____ or IRB App. Date _____ <input type="checkbox"/> INTERNATIONAL ACTIVITIES: COUNTRY/COUNTRIES INVOLVED (GPG II.C.2.j) _____ <input checked="" type="checkbox"/> COLLABORATIVE STATUS <b>Not a collaborative proposal</b>					

# *Campus Cyberinfrastructure*

A Coordinated Investment by Many Stakeholders

NSF and other agencies

Campuses

CIOs

VPRs

Faculty

*My talk will focus mostly on the local stakeholders!*



# *Communicating Value*

## Hot Take: Our Community is Immature at Communicating Value



Questions we received 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?”

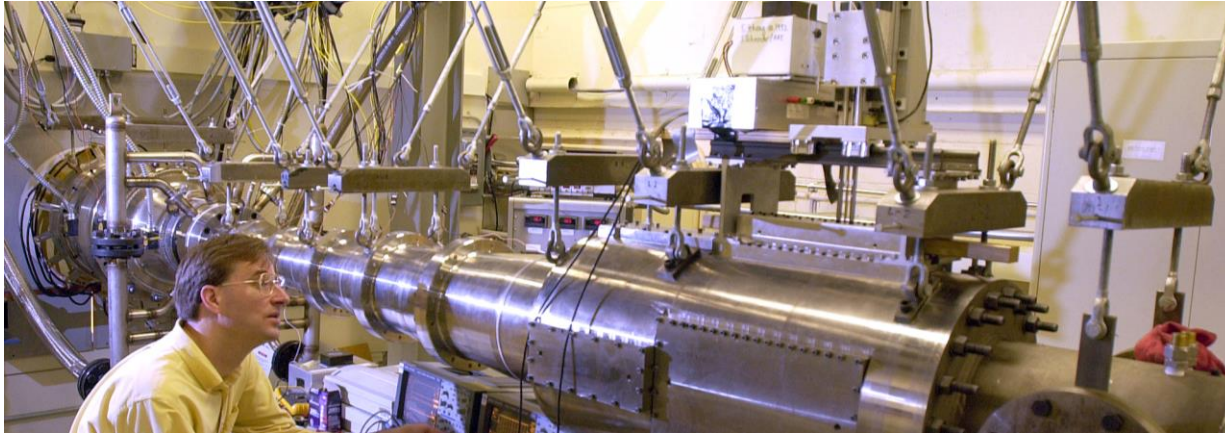
“Why is that recharge center not covering its costs?”

Which I understand.. Even if you do recover all the costs, the **VALUE** is returned somewhere else than your center’s books.



# 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?

## *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, happy faculty

**Provost:** Students,  
educational outcomes,  
happy faculty

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

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

**CIO:** Contributing to the  
mission

## *Personas*

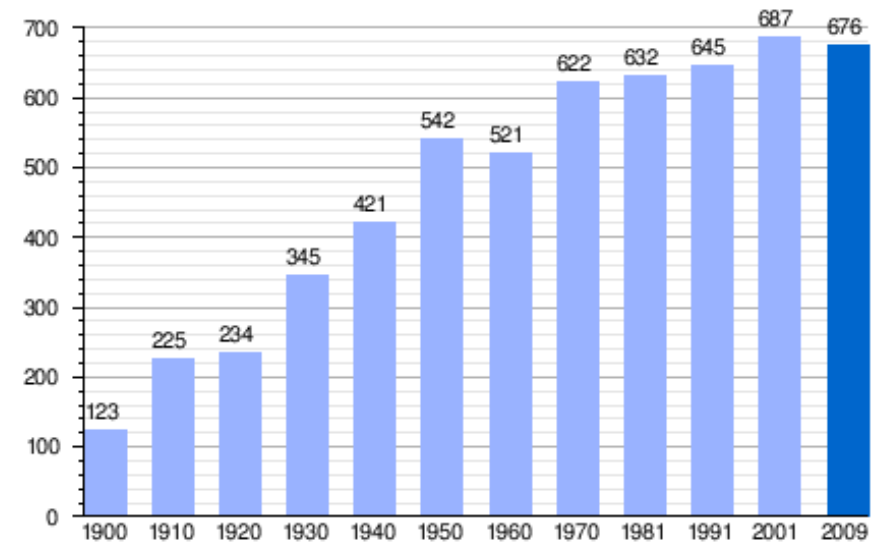
Stakeholder Interests

## *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

If you don't track your usage, and map it to who your PIs are and where they sit in your institution, start doing so now! (Install XDMOD!)





# *Purdue, circa 2001*

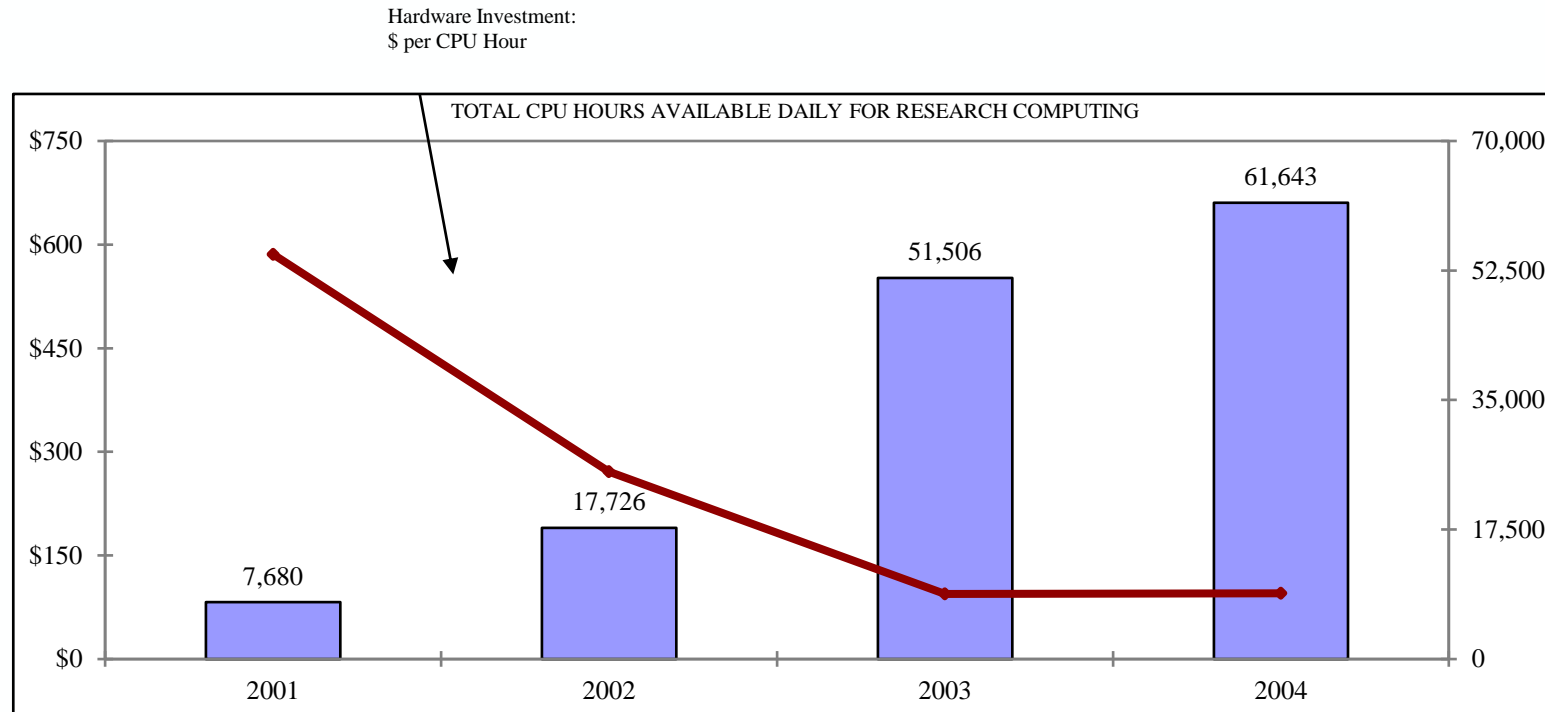
Office of the VP for IT formed, first CIO



- Jim Bottum hired from NCSA as first CIO
- Administrative IT and computing center merged, forming "IT at Purdue" (ITaP)
- Prior to hiring of CIO, computing and storage capacity was behind where it should be, newest compute resource was an 8-year old IBM SP
- Getting research computing in order was a top priority
- Got into the "Beowulf" world, and started our "Community Cluster" condo program ~2004 or so

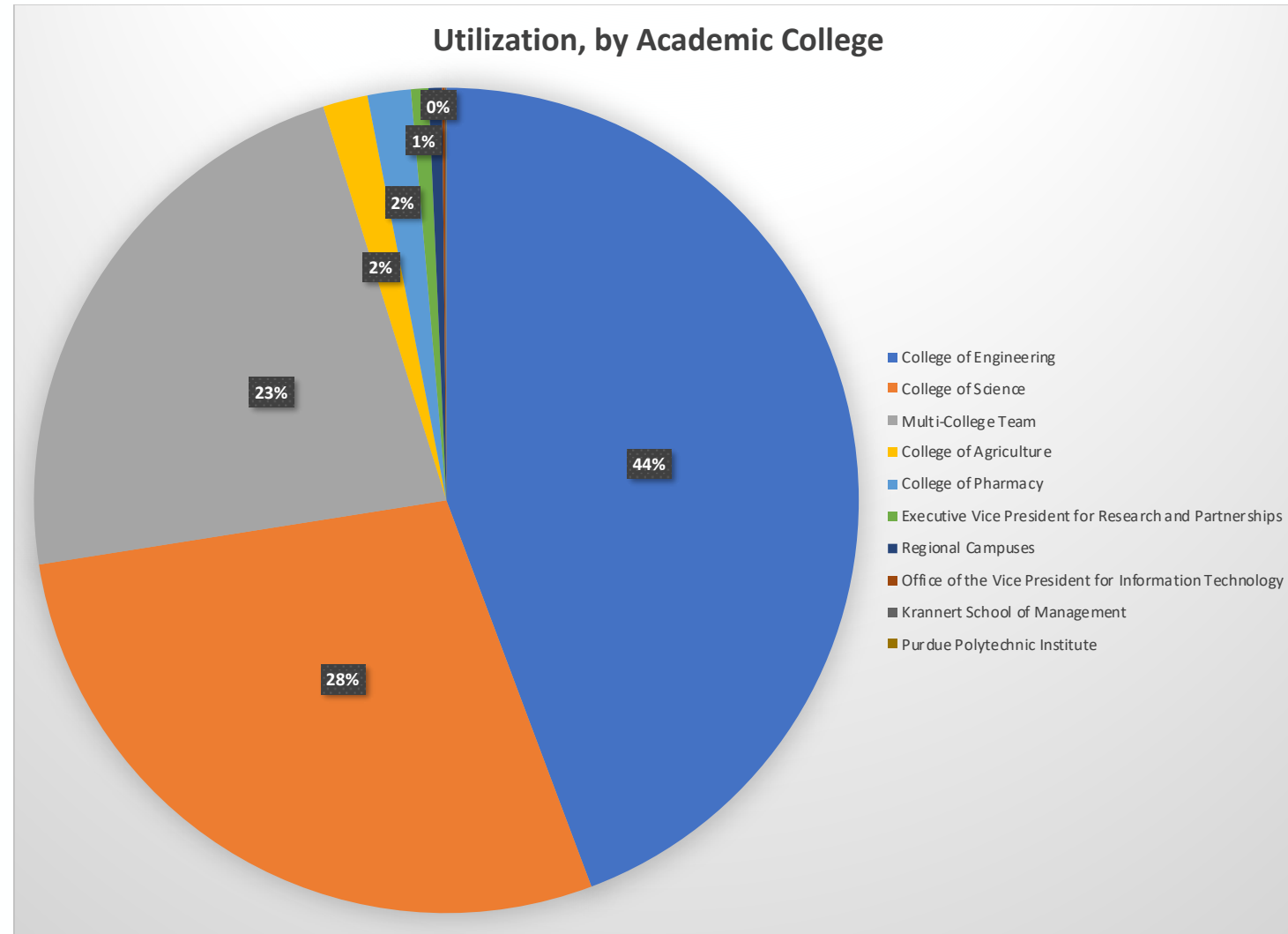
# *To All Stakeholders:*

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



# *To Faculty, Academic Leaders:*

Who is Using the Cyberinfrastructure?



# *Are You Cost-Effective?*

Variations of these questions include:

- Why don't you use the cloud?
- Didn't you **just** buy an HPC system?
- How is this different than the other thing you just got?
  - ... or that INSERT PEER HERE just got?

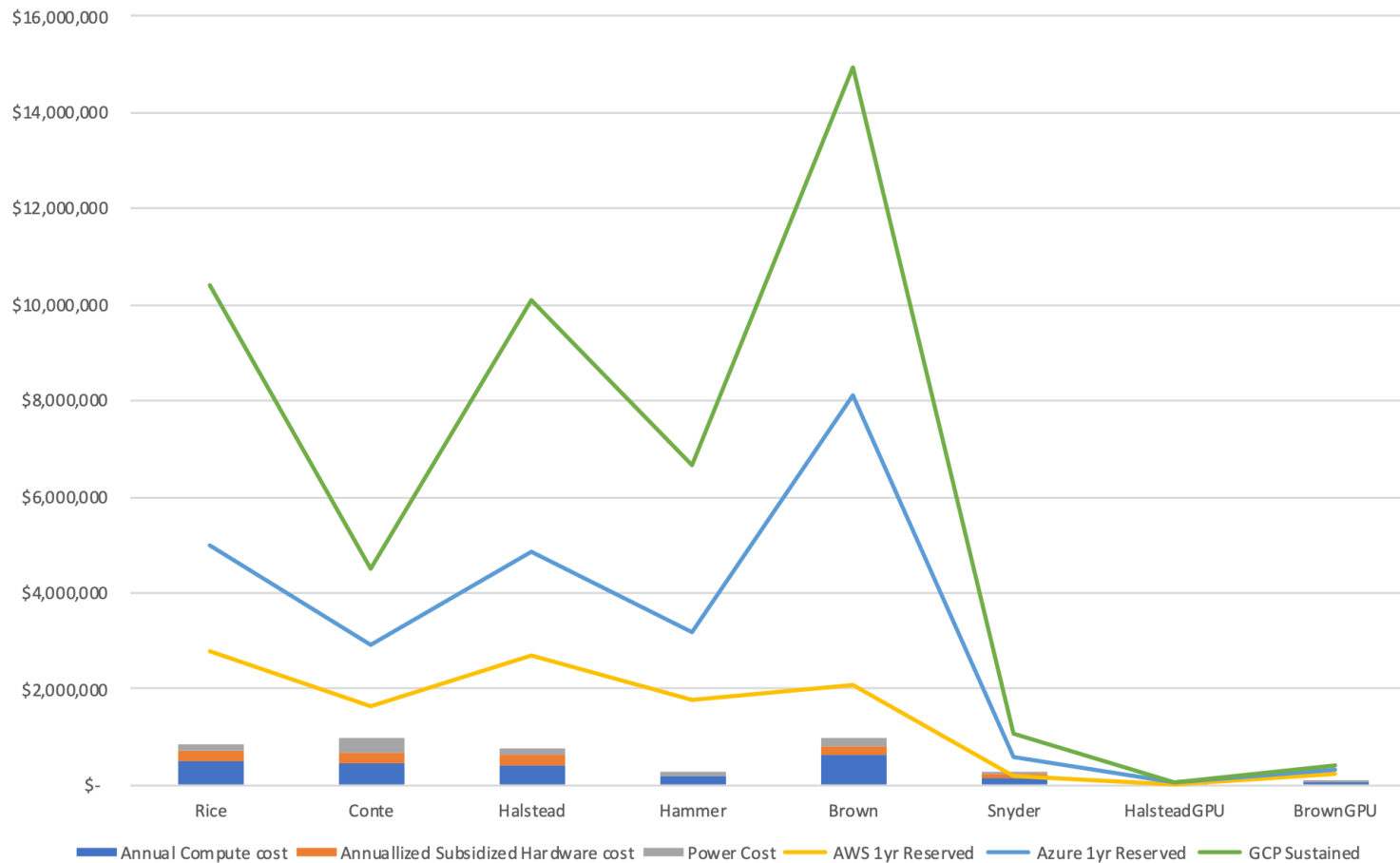


<https://www.flickr.com/photos/85567416@N03/14224208945>

# *To CFO and Faculty:*

Our model is more cost effective than commercial cloud (2019)

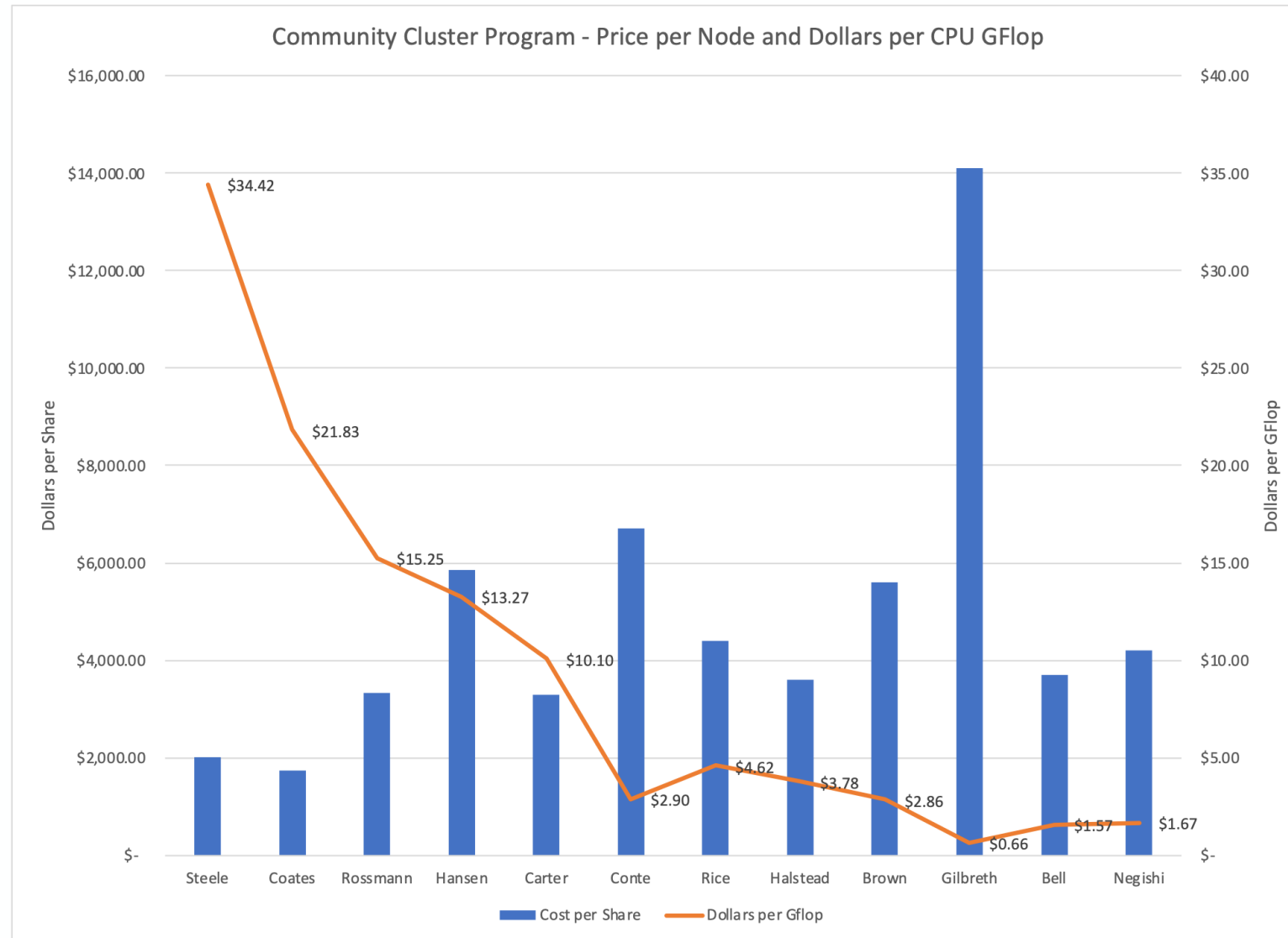
Annual Cost - 2018 Per-System Community Cluster Usage on  
Public Clouds vs On-Prem  
(Including HW Subsidy)





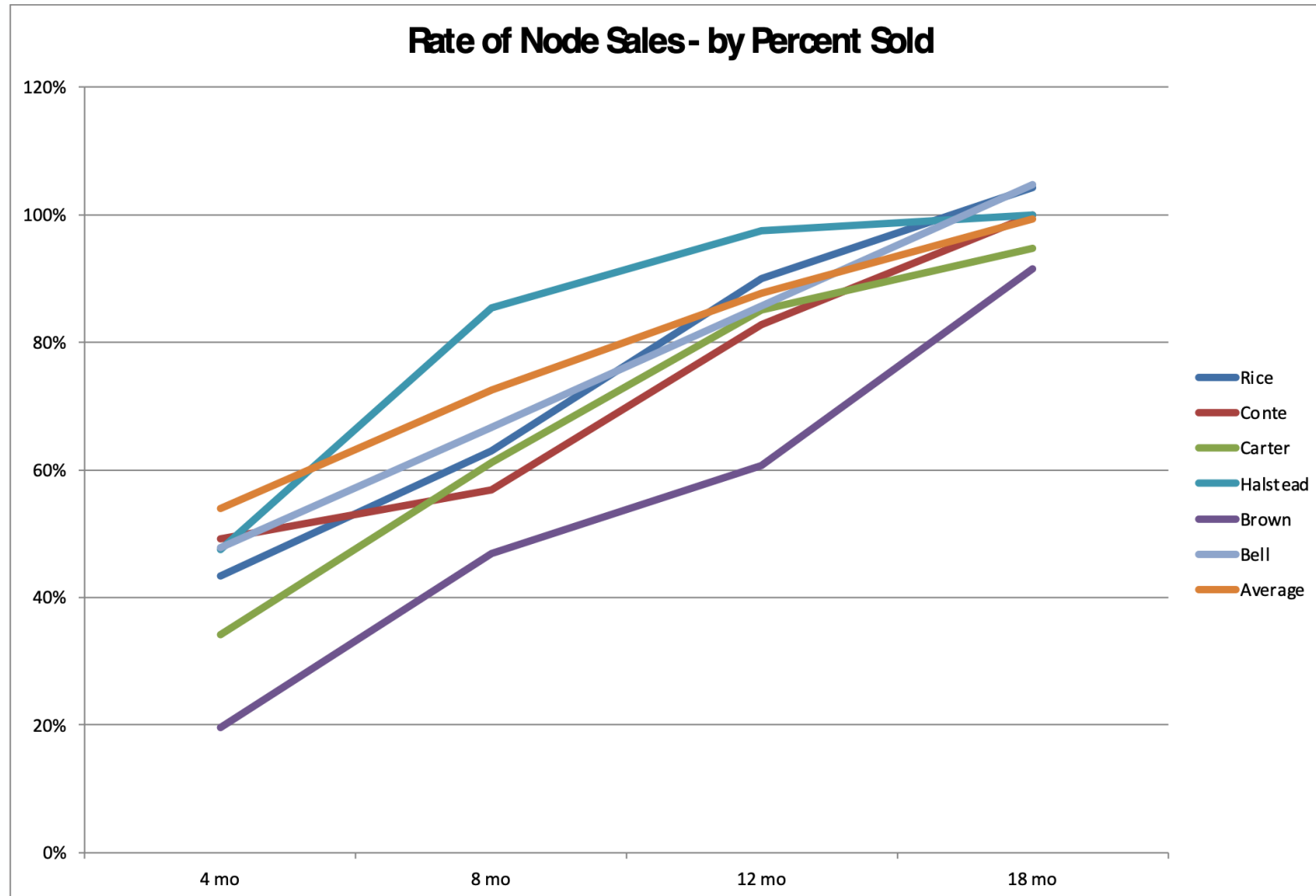
# *To CFO, and Faculty:*

Each iteration gets us more capability per dollar



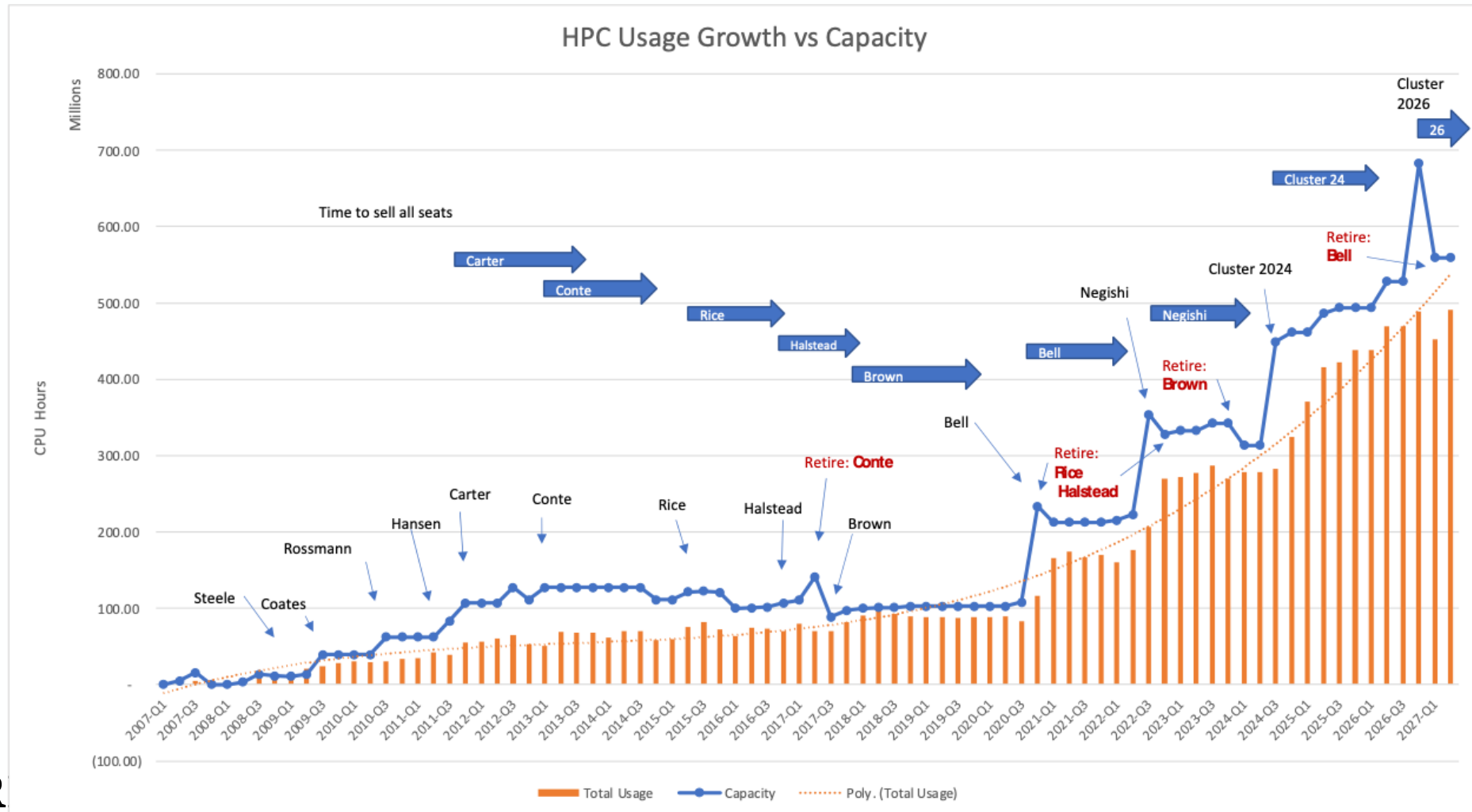
# *To All, but mostly CIO and CFO:*

We understand our sales data well enough to plan our lifecycle needs



# *To CIO and CFO:*

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



# *Are You a Partner?*

From Purdue's 2002 IT Strategic Plan

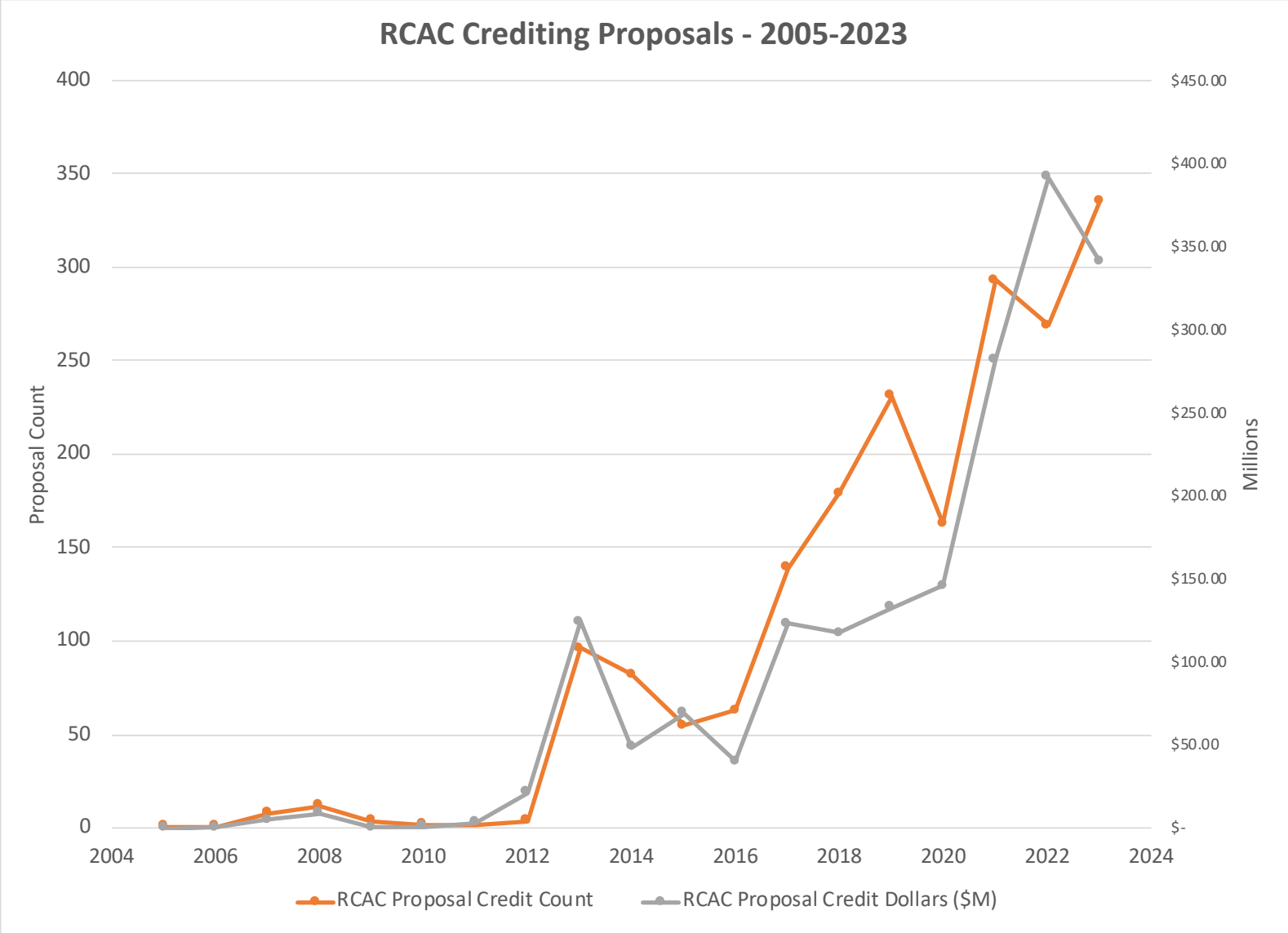
- Project has to be part of University's strategic direction
  - Work out priorities with the VPR
- Extra points for interdisciplinary projects
  - Consistent with Purdue strategic directions
- Funding potential
  - Two approaches to proposals
- Both sides have something to gain (in each other's critical path)
  - In addition to monetary
- Project should be innovative
- Write it down (project management)
- Resources (have done loss leaders but...)
  - Loss leaders early to establish credibility



# *To Faculty and Academic Leaders*

Our partners cite our center as necessary to their proposals

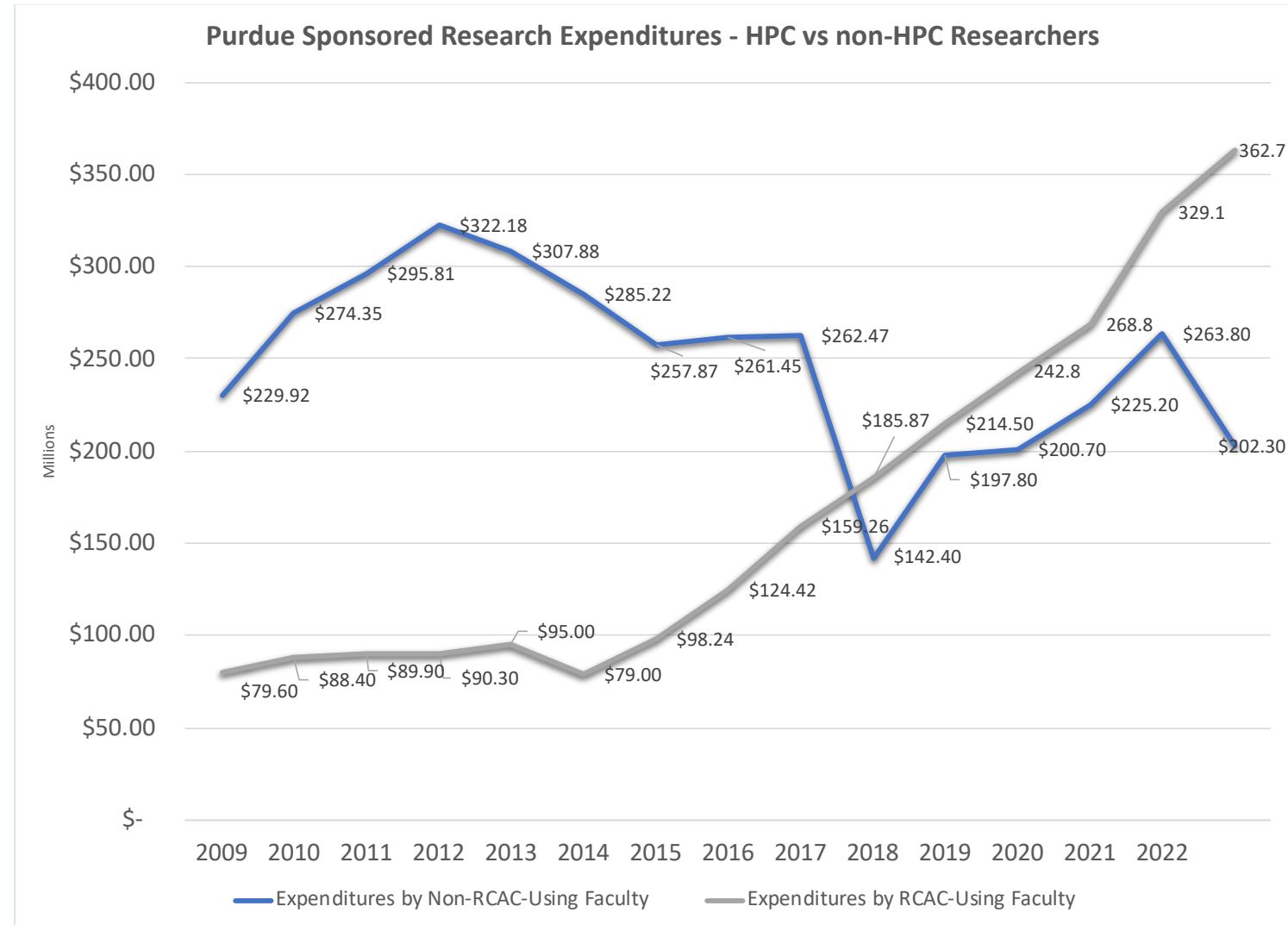
RCAC Crediting Proposals - 2005-2023





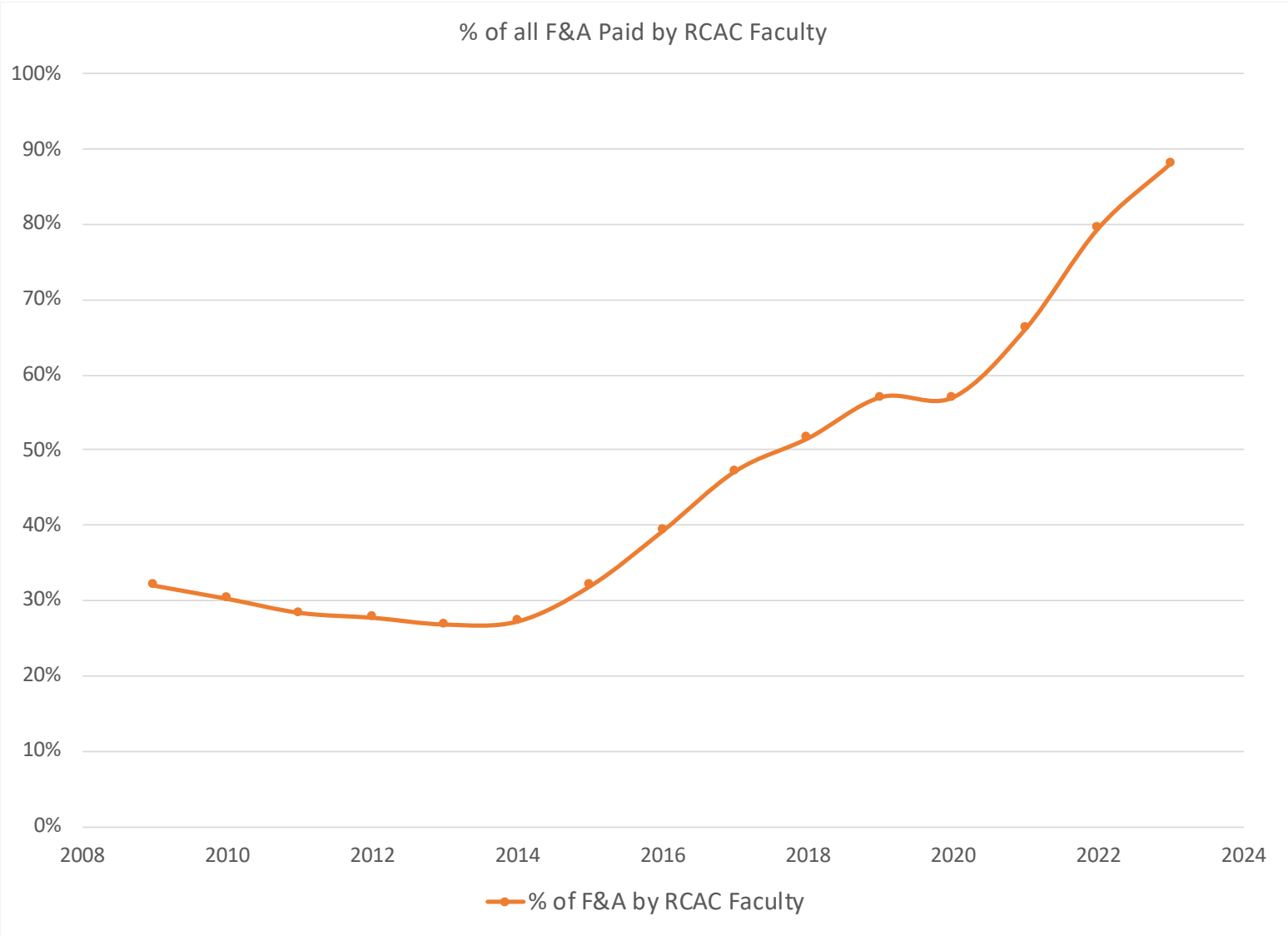
# *To Faculty, Academic Leaders, CFO*

An increasingly large proportion of the research dollars go to users of HPC



# *To the CFO*

Users of our center pay most of the university's F&A



*To all:*

Institutional investment in research scientists and RSEs has a positive ROI

<b>FY</b>	<b>Scientist and RSE General Fund Salaries (\$K)</b>	<b>Research Expenditures (\$K)</b>	<b>F&amp;A Paid (\$K)</b>	<b>ROI to General Fund</b>
2020	\$226.00	\$2,900.00	\$926.00	4.10
2021	\$215.40	\$2,710.00	\$747.30	3.47
2022	\$108.10	\$13,500.00	\$730.00	6.75
2023	\$73.60	\$3,400.00	\$986.00	13.40

# *Are your Metrics Tied to a Product?*

## Sage Advice

- Last CIO: *“Raw count metrics are meaningless to the administration unless related to the business of Purdue. You need to talk about the value in those terms”*

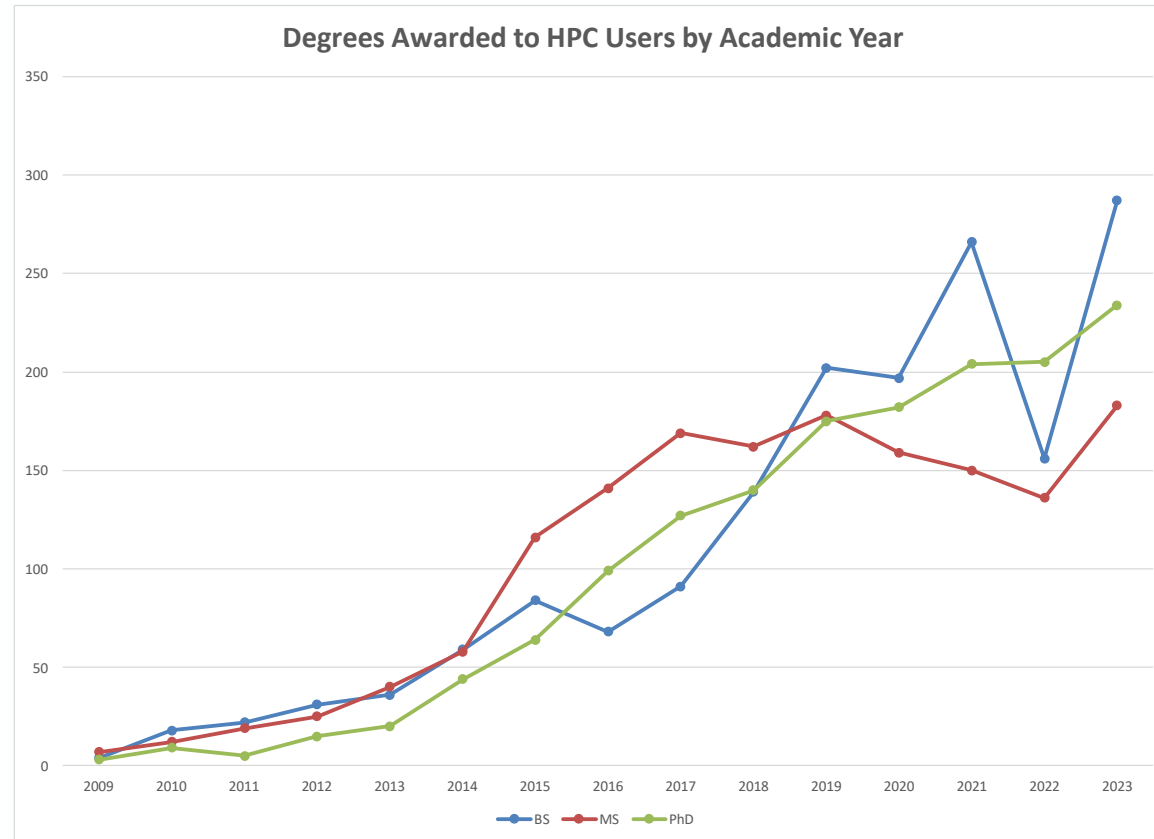
### That Administration's Priorities:

- Control Costs (tuition freeze)
- Impact Students
- Recruit and retain faculty



# *To Faculty and Academic Leaders*

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



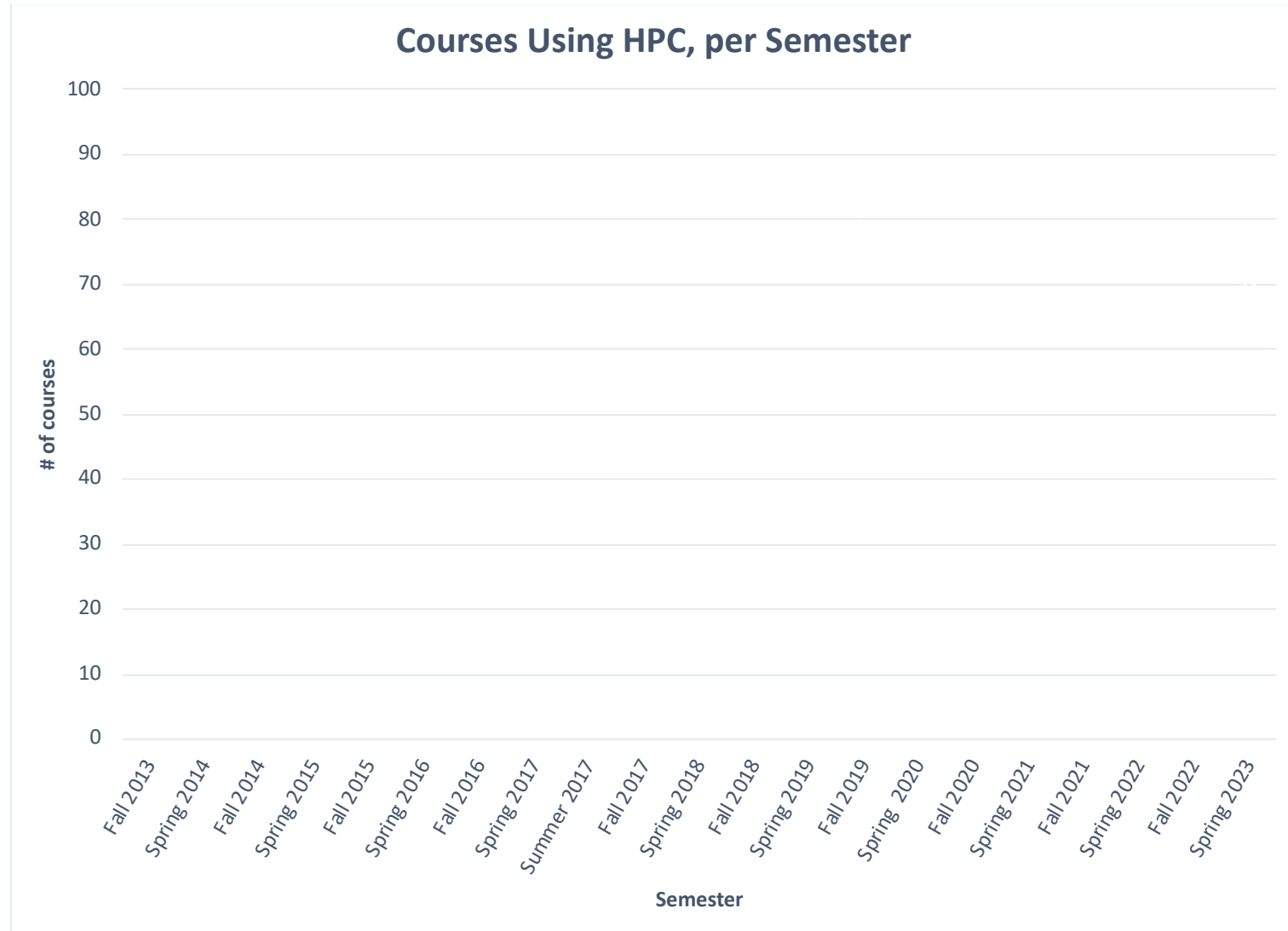
Year	Earned Doctorates	HPC-Using Doctorates	% Using HPC
2010	639	9	1%
2011	672	5	1%
2012	656	15	2%
2013	687	20	3%
2014	735	44	6%
2015	709	64	9%
2016	727	99	14%
2017	740	127	17%
2018	758	140	18%
2019	738	175	24%
2020	808	182	23%
2021	802	204	25%
2022	835	205	25%
2023	851	234	27%



# *To Faculty and Academic Leaders*

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

- During the 2022-23 AY, 167 courses used HPC for instructional purposes, impacting a total of over 3000 students.



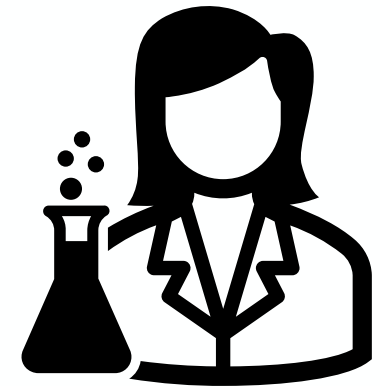
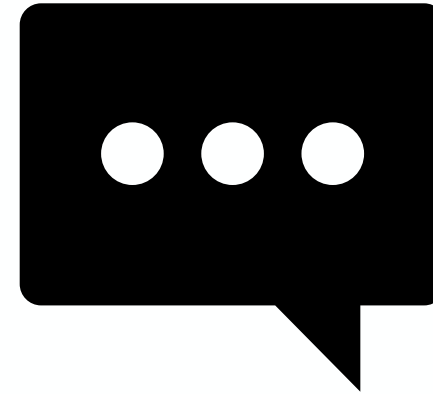
# *To the CFO*

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

2023 Metric	Amount (\$M)
Net Cost to Purdue	\$7.55
Awards to RCAC-using Faculty	\$395.00
Direct Expenditures Enabled	\$362.70
F&A Expenditures Enabled	\$77.50
Return on Investment	
Awards Return	52.32
Direct Expenditure Return	48.04
F&A Expenditure Return	10.26

## *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.”



# Testimonials



## ***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



## ***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



# *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

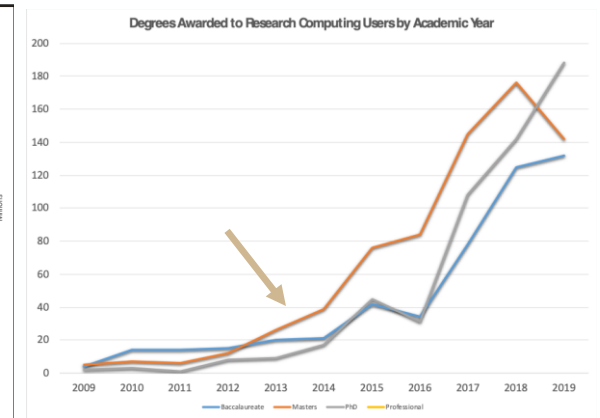
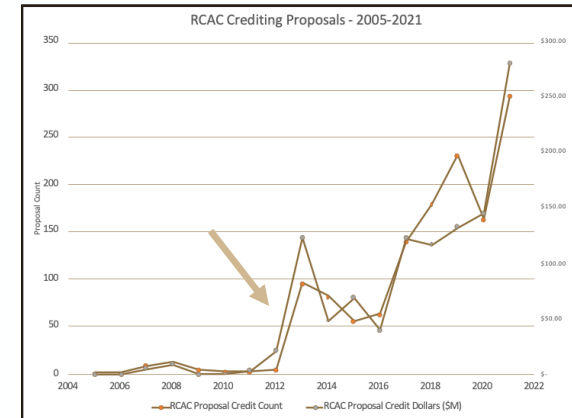


Purdue Polytechnic Institute  
West Lafayette, Indiana  
August 2022

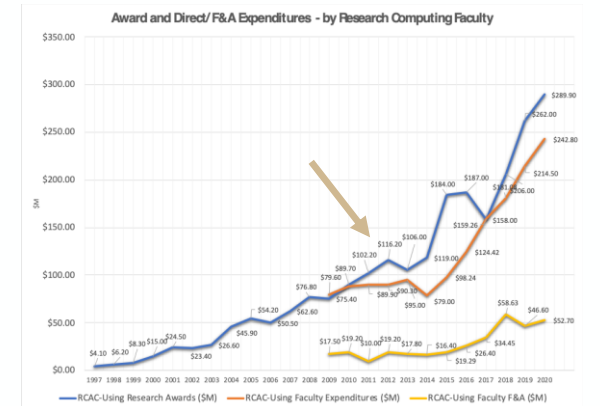
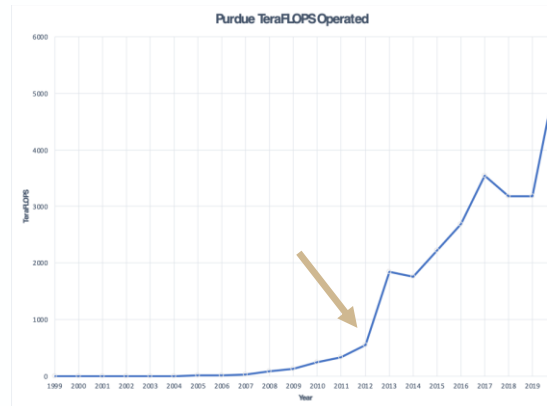
# Motivator: Metrics at Purdue

- 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!*



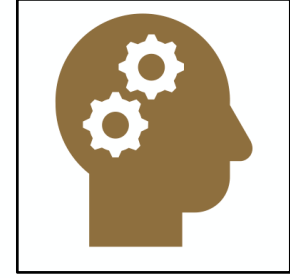
# *Research Questions*



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



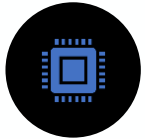
- If so, what are the key factors for investment that lead to returns on institutional outputs?



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

# 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.

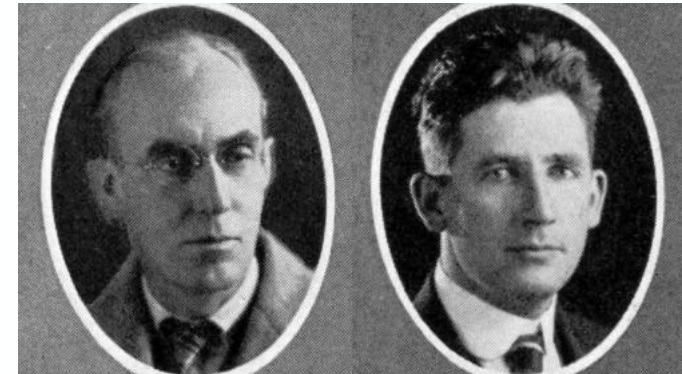
- **Downsides:** Top 500 is no longer a viable proxy for investment.
- How to add a labor dimension in model? No cross-institutional data available as to investment.

# ***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)*

# *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

# Correlation Analysis

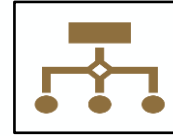
For Single Inputs TF and Salary

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

# *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

.. And another series of regressions to model  
the capability of the labor variable, using  
research expenditures by center staff as a  
metric.



# *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%

# *Output – Earned Doctorates*

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

# *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

# *Broader Applications*

How can this framework be applied to other CIs?

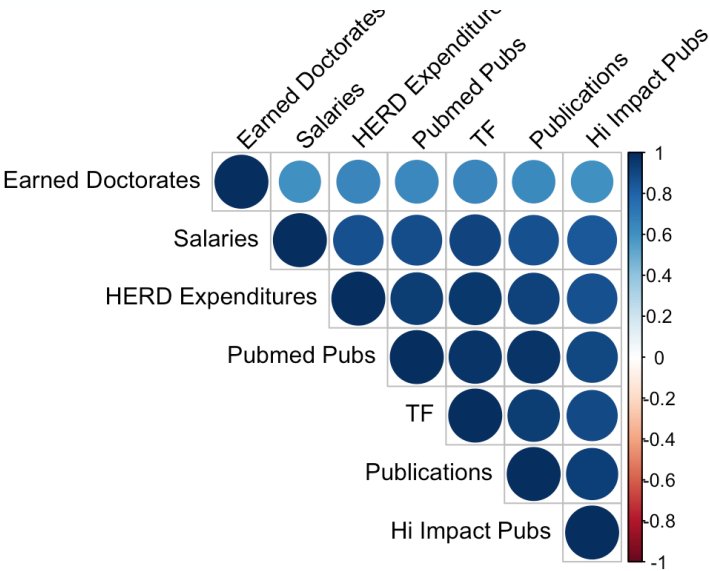
- Work ongoing to apply model to other campus' data.
  - Some friendly institutions have shared
    - *Can I talk to you about your data?*
- Improve the model – use timeseries statistics
- Apply model to NSF XSEDE infrastructure
  - Using all NSF expenditures, US Earned Doctorates, NSF-funded pubs in Scopus, etc.

# Institution 1

term<chr>	TF<dbl>
Pubmed Pubs	0.9736463
HERD Expenditures	0.9615496
Publications	0.9448686
Salaries	0.9117805
Hi Impact Pubs	0.8969057
Earned Doctorates	0.6570913

6 rows

	HERD Exp (\$M)	Pubs	Hi Impact Pubs	PhDs
100 TeraFLOPS	10.11	10.11	0.88	0.65
\$100k Salaries	11.19	109.90	5.75	4.68

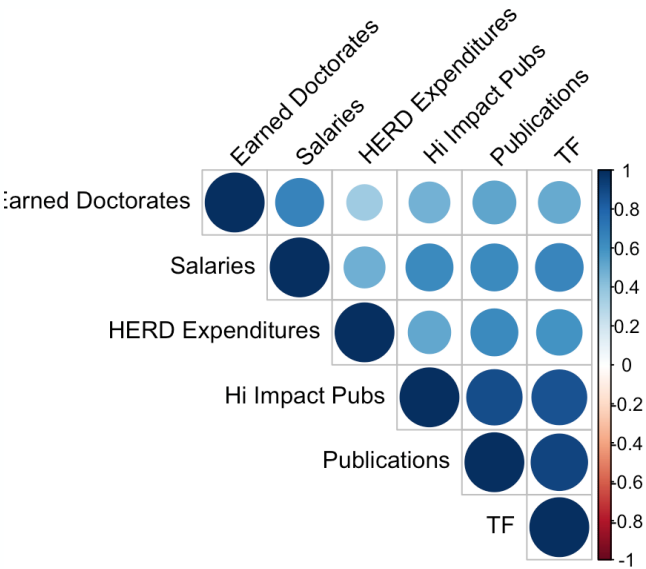


# Institution 2

term<chr>	Salaries<dbl>
Earned Doctorates	0.6610407
TF	0.6522726
Publications	0.6329621
Hi Impact Pubs	0.6329621
HERD Expenditures	0.4810512

5 rows

	HERD Exp (\$M)	Pubs	Hi Impact Pubs	PhDs
100 TeraFLOPS	25.50	25.50	5.17	-0.29
\$100k Salaries	15.61	86.04	13.36	3.89



# *Thank You*

psmith@purdue.edu

