A Practical Toolkit

For Demonstrating Return on Investment

In Cyberinfrastructure

Quantify • Qualify • Communicate

Prepared By: Esen Gokpinar Shelton, Ph.D.[1]; Winona Snapp-Childs, Ph.D.[1]; Preston Smith, Ph.D.[2]

[1]Indiana University, [2]Purdue University

Why

ROI in CI Matters?

Clusters are expensive		Leaders and Funders ask:
Staff, hardware, storage, electricity, networking.	**+**	Is it worth it? Value often remains opaque.

Without ROI → risk of underfunding, leadership questioning, or losing talent.

ROI stories help you:
- Secure additional investment (e.g., renewal, scaling).
- Justify cost-sharing requirements (CC* STAR requires 30% match).
- Demonstrate responsible stewardship of funds.

The CHALLENGE:

Unlike a business, ROI isn't measured in profit for CI in higher education.

How to Measure Value Beyond Profits?

Prior Studies Show...

Research computing benefits:

- Increased research productivity
- Enhanced training opportunities
- Workforce development

However ...

- Few provide a practical, step-by-step toolkit for demonstrating ROI
- Focus often on economic methodologies, storytelling, or funding models.

There is a NEED for an actionable framework that practitioners can readily apply, and this project aims to fill that gap.

This Toolkit is for

Three Primary Audiences

Institutions maintaining existing clusters and considering expansions or refresh cycles

(who must show the tangible and intangible value of prior investments)

Applicants for awards like NSF CC*

(who face requirements to demonstrate cost share (typically ~30%) and to make the case for sustainable support)

Faculty champions and campus leaders advocating for CI with provosts, vice presidents, deans, and trustees

(who need evidence that resonates beyond technical metrics)
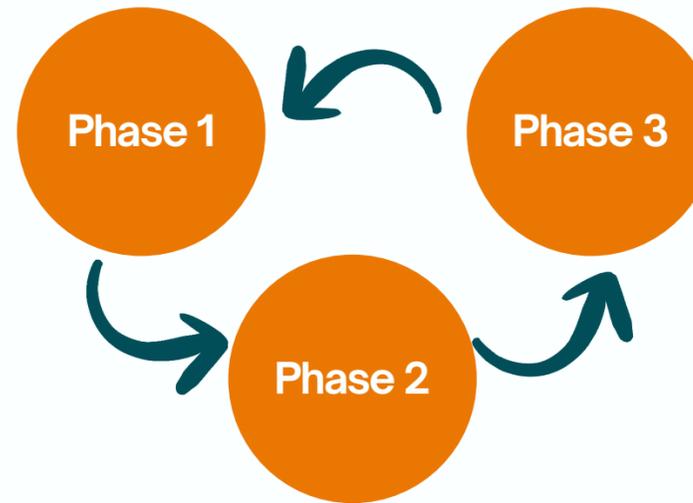
Our Approach

Mixed-Methods, Stakeholder-Centered

- **Quantitative metrics**
  - Grants, utilization, efficiency

- **Qualitative evidence**
  - Researcher stories, workforce development, societal impact

- **Iterative framing**
  - ROI as an ongoing practice

Toolkit Framework

3 Phases, 8 Steps

Phase 1

Phase 3

Phase 2

The process is not linear, it is itiretative.

Phase 1: Foundations

Step 1	Scope & Purpose
Step 2	Stakeholders & Value Dimensions
Step 3	Baseline & Counterfactuals

Phase 2: Measurement & Valuation

| Step 4 | Quantitative Indicators |
| Step 5 | Qualitative Evidence |

Phase 3: Communication & Governance

Step 6	Dissemination
Step 7	Engagement
Step 8	Systematize/ Standardize

Phase 1:

Foundations

The **_Foundations_** phase is critical because it sets the stage for meaningful ROI assessment.

If purpose, stakeholders, and baseline context are not clearly defined, ROI calculations can become confusing, misaligned, or unpersuasive.

Foundations are about asking the right questions, understanding the context, and clearly documenting what the ROI effort will address. This phase consists of three steps.

Step 1 Scope & Purpose

Step 2 Stakeholders & Value Dimensions

Step 3 Baseline & Counterfactuals

Define Purpose & Scope

Before collecting data, clarify *why* the ROI assessment is being conducted and *what* it will cover.

Without a clear purpose, efforts risk unfocused data, missed metrics, or reports that fail to resonate with decision-makers.

## Key Framing Questions: Always ASK!

- **What decision or opportunity is driving this analysis**?
  - Cluster refresh or expansion?
  - NSF CC* proposal (cost-share)?
  - Faculty or staff advocacy for new resources?

  ....

- **What timeframe is meaningful**? (3, 5, or 10 years depending on context)

  ....

- **What resources should be included?** (Hardware only, or storage, software, staff, training, support)

Example Scenario #1

Cluster Refresh

Imagine you are a research computing professional, helping campus leaders decide whether to upgrade or replace a compute cluster.
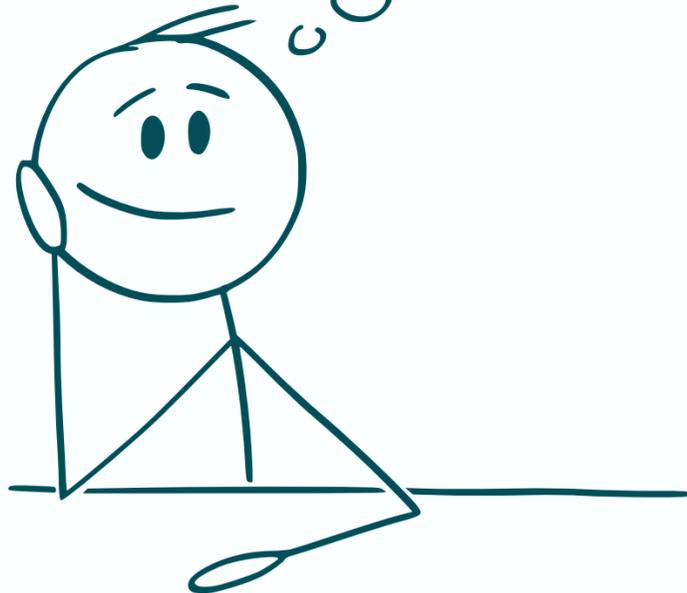
What Should You Do?

1. **Define your scope**: Focus on the previous system's 5–10 year lifecycle. This allows you to capture meaningful trends such as utilization patterns, job completion rates, and research output.

2. **Align with decision-making:** A timeframe that matches budget cycles and depreciation schedules makes it easier for CFOs and provosts to understand the financial logic.

3. **Build your ROI story:** Present clear, measurable outcomes.

"A $500K investment supported $8M in grant funding and significantly reduced compute wait times over five years "

Such statements give leadership tangible evidence to justify a refresh.

Example Scenario #2

NSF CC* Proposal

Imagine you are a research computing professional preparing an NSF CC proposal.*

What Should You Do?

1. **Define your scope**: Demonstrate alignment with NSF's solicitation, not just local decision-making.

2. **Expand your scope**: Include hardware, training, partnerships, and workforce development programs to show the full institutional impact.

3. **Build your ROI story:** Document measurable outcomes such as:
   - Cluster utilization
   - Number of students trained in workshops
   - Collaborations with regional partners
   - Workforce placements

Why This Matters:
Framing ROI broadly ties institutional investments to NSF priorities, strengthens cost-share arguments, and ensures your analysis is proposal-ready for funders.

Example Scenario #3

Faculty Advocacy

Imagine you are a principal investigator advocating for additional campus or college-level resources.

What Should You Do?

1 **Define your scope**: Demonstrate tangible benefits within a single department or college, rather than the whole institution.

2 **Focus your scope**: Highlight local impacts such as:
- Number of graduate theses supported
- Publications produced using cluster access

3 **Build your ROI story:** By starting with concrete, localized results, you provide clear evidence that resonates with departmental or college leadership.

Why This Matters:
A focused, incremental approach generates early buy-in and can later be expanded into an institution-wide ROI narrative, especially in settings where central support for research computing is still developing or contested.

In Short....

Purpose & Framing Differ by Context

Key Takeaway:
- ROI framing should adapt accordingly

Scenario	Purpose	How to Frame ROI	Example
Cluster Refresh	Justify a new or upgraded system	Look back at the previous lifecycle (5-10 years): Utilization, job completion, research output, financial return	500K cluster → $8M in grants + reduced wait times → leadership approves refresh
NSF CC*	Align with NSF priorities	Broaden scope beyond hardware: Include training, partnerships, workforce development	Document cluster use + students trained + regional collaborations → strengthens cost-share case
Faculty Advocacy	Secure local/departmental support	Start small, show tangible benefits to one unit or discipline	Evidence: theses completed, publications enabled → builds early buy-in for wider support

Identify Stakeholders & Value Dimensions

Identify the stakeholders who will engage with your ROI assessment and to **understand what "value" means in their terms**.

Even the most carefully collected data will fail to persuade if it does not speak the language of the intended audience.

Different stakeholders, different priorities!

- **Senior Admins** (provosts, vice presidents, deans)
  - Financial leverage, institutional reputation, and research competitiveness.
- **Faculty**
  - How cyberinfrastructure accelerates results, strengthens grant proposals and collaboration.
- **Students**
  - Workforce skills, employability, and access to advanced tools.
- **Funders**
  - Sustainability, societal impact.

Value Itself is also Multifaceted

Drawing on frameworks such as **Integrated Reporting**, value can be thought of across several dimensions:

Financial
Grants, cost savings, operational efficiency.

Intellectual
Publications, software, patents, innovations

Human
Students trained, doctoral completions, workforce pipelines

Societal/Relational
Collaborations, partnerships, community engagement

Natural
Environmental contributions, energy efficiency, sustainability outcomes

Stakeholders x Value Dimensions

The key is not to treat the dimensions as abstract categories, but to map them to the right stakeholders.

This exercise keeps the ROI story from being too generic. By tailoring evidence to the concerns of specific groups, the resulting analysis becomes not only rigorous but also persuasive.

Stakeholder	Financial	Intellectual	Human	Social/Relational	Natural
Provost/VP	X	X			X
Faculty		X	X		
Students			x		
Funders	X	X	X	X	X

Establish Baseline & Counterfactuals

ROI is most convincing when it answers the question:

What would have happened without this investment?

Two Key Elements:

- **Baseline** – The state before the investment
  - Long compute queues delaying research
  - Faculty paying out-of-pocket for commercial cloud
  - Fewer students trained due to limited resources

- **Counterfactual** – What if the investment hadn't been made?
  - Higher costs for commercial cloud services
  - Lost competitiveness in national awards
  - Missed workforce training opportunities

Examples in Action

Campus Wide Efficiency

Outsourcing cluster jobs to commercial cloud would have cost 3x more annually → **avoided costs highlighted to provost.**

Research Competitiveness

Genomics PI showed NIH renewal would have cost $250K more on AWS, with delayed results → **campus computing framed as critical to staying competitive.**

Student Training

Without the cluster, graduates would lack exposure to real-world-scale datasets → **dean showed link to employability & institutional reputation.**

Practice Time!

Use the following link to a worksheet to apply the three **Foundations** steps to your institutional context. Jot down brief notes or bullet points for each section. Don't worry about getting it perfect, this is about sharpening your thinking.

NOW, ITS YOUR TURN

Link:
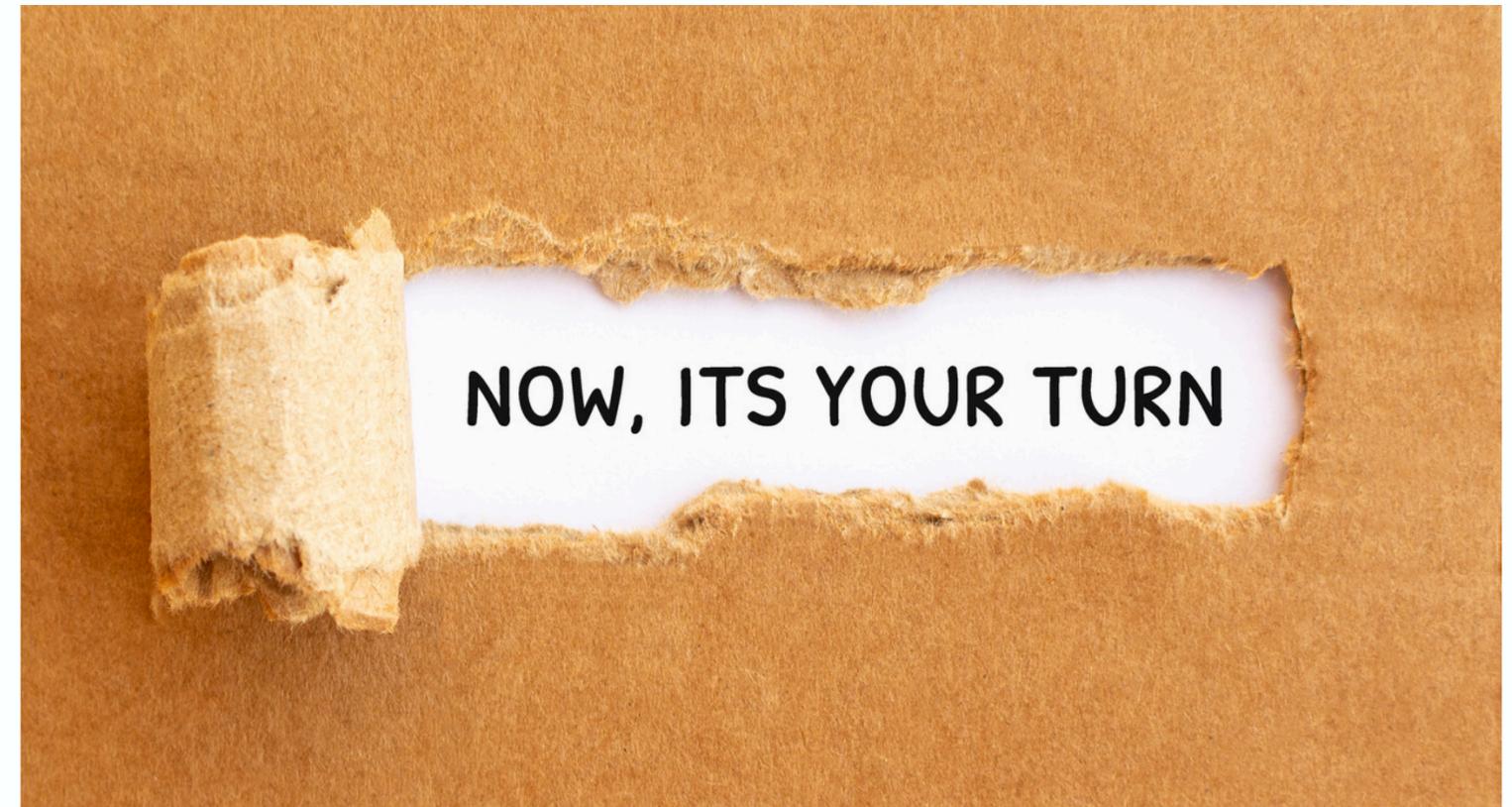
Phase 2: Measurement & Valuation

Move from Framing → Evidence Gathering

- Measurement isn't just about collecting numbers, it's about making CI impact:

  1 Visible
  2 Persuasive
  3 Understandable

Strong ROI cases combine:

- Numbers → provide credibility, benchmarking, scale
- Stories → connect metrics to people, competitiveness, and reputation

| Step 4 | Quantitative Indicators |
| Step 5 | Qualitative Evidence |

Collect Quantitative Indicators

Quantitative evidence = ROI backbone.
Focus on what to measure + how to gather:

Productivity Outputs

- *Publications, patents, datasets, software*
- *Example: R1 = hundreds of papers, high citations;*
- *R2 = "5 major papers + 1 software tool that put us on the map"*

Financial Leverage

- *Formula: Grant funding supported ÷ CI investment*
- *Example: $500K cluster → $8M grants over 5 yrs → 16× leverage*

Human-Capital Development

- Students trained, workshops, certificates
- Example: 3,500 trained → half placed in labs/industry.

Utilization & Access

- *Metrics: core hours, jobs run, users*
- *Example: R1 = millions of jobs across dozens of departments; Small school = 200 undergrads using cluster in coursework.*

Operational Savings

- *Compare local costs vs. cloud (AWS/Azure)*
- *Example: Genomics team avoided $225K in cloud costs.*

- $500K cluster → $8M grants, 3,000 students trained, 120 publications

- Case: Genomics project ran locally for $25K (vs. $250K on cloud), enabling timely NIH submissions

Not Every Institution
Can Track All Five Categories with Equal Rigor

- R1s: rich data infrastructure → can capture scale (bibliometrics, admin records).
- R2s & smaller schools: rely on scheduler logs, PI self-reports, student course data.

Best practice: Pick 2–3 categories that:

- Are feasible to gather
- Resonate with local decision-makers
- Can expand over time
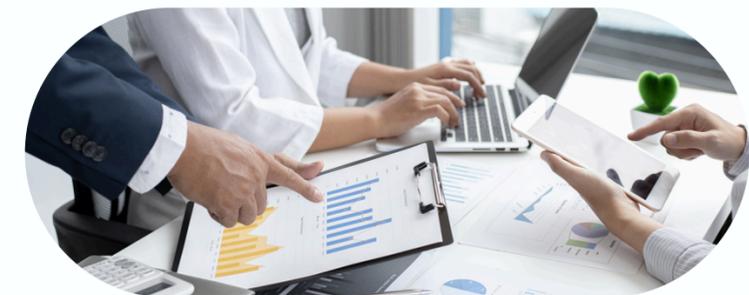
Capture Qualitative Evidence

- Numbers show scale, but stories make impact stick.
- Provosts, trustees, and funders often remember narratives, not raw data.
- Qualitative evidence adds a human dimension that complements quantitative indicators.

Examples

- Researcher experiences → "CI let us publish 6 months faster."
- Student trajectories → "From workshop to data scientist at NIH."
- Societal impacts → "Without this cluster, hurricane forecasts for 2022 would have failed."

Sources of Qualitative Evidence

- **Researcher testimonials**
  - New discoveries, faster results, competitiveness.
- **Case studies**
  - CI role in pressing issues (e.g., pandemic modeling, climate forecasts).
- **Student stories**
  - Training → workforce placement (e.g., Oak Ridge, Google).
- **Societal outcomes**
  - Health, sustainability, national security.

Example:
- "Without these systems, we would have been blind to critical dynamics of the pandemic." – COVID-19 PI using XSEDE

Pairing Numbers with Stories

- The formula for persuasion:
  - Metric → credibility, scale
  - Story → meaning, memorability

Examples:

Metric: 3,500 students trained

Story: "Jane Atkinson discovered HPC here; now she's at NIH."

Metric: $8M in grants supported

Story: "The cluster was the difference between being competitive and shut out of NIH funding."

Practice Time!

Use the following link to a worksheet to apply the Phase 2: Measurement & Valuation steps to your institutional context. Jot down brief notes or bullet points for each section. Don't worry about getting it perfect, this is about sharpening your thinking.

Link:

NOW, ITS YOUR TURN

Phase 3:

Communication & Governance

Phase 3 focuses on making ROI visible, meaningful, and sustainable.

At this stage, **the emphasis shifts from analysis to strategic communication, targeted dissemination, and embedding ROI practices into institutional routines.**

Without effective communication and governance, even the most rigorous evidence may fail to influence decisions or secure continued support

Step 6	Dissemination
Step 7	Engagement
Step 8	Institutionalization

Disseminate ROI Findings

- Evidence only matters if it's communicated effectively.

- **Different Audiences = Different Framing.**

Format Strategically:

- **Provosts/Admins**:
  - 1-page briefs or dashboards → ROI multiples, grants, training highlights
- **Trustees/Boards**:
  - Slide decks → workforce, financial leverage + 1–2 case studies
- **Faculty/Internal Groups**:
  - Story-driven talks → operational impact, acceleration, mentorship
- **Funders (NSF/NIH):**
  - Detailed reports and appendices → full methodology + societal/scientific outcomes

Engage Stakeholders Continuously

ROI communication must be ongoing, not one-off.

Engagement = Dialogue, not broadcast. ROI story should evolve with shifting priorities.

- **Recurring Updates** → Annual/biannual ROI reports, dashboards, newsletters

- **Interactive Forums** → Workshops & briefings where faculty, students, and admins review results + share case studies

- **Feedback Loops** → Gather input on what resonates (e.g., trustees → workforce, boards → sustainability) and refine future reporting

Stystematize ROI Practices

ROI must be routine, transparent, and embedded in governance to sustain impact.

Key Strategies

- **Regular Reporting Cycles** → Annual ROI reports/dashboards tied to planning & budget reviews
- **Shared Ownership** → RCD + research offices + admin units maintain data & storytelling
- **Standardized Methods** → Consistent metrics/calculations enable year-over-year comparison & benchmarking
- **Link to Decisions** → ROI reporting feeds directly into staffing, upgrades, funding proposals

Practice Time!

Use the following link to a worksheet to apply the Phase 3: Communication & Governance steps to your institutional context. Jot down brief notes or bullet points for each section. Don't worry about getting it perfect, this is about sharpening your thinking.
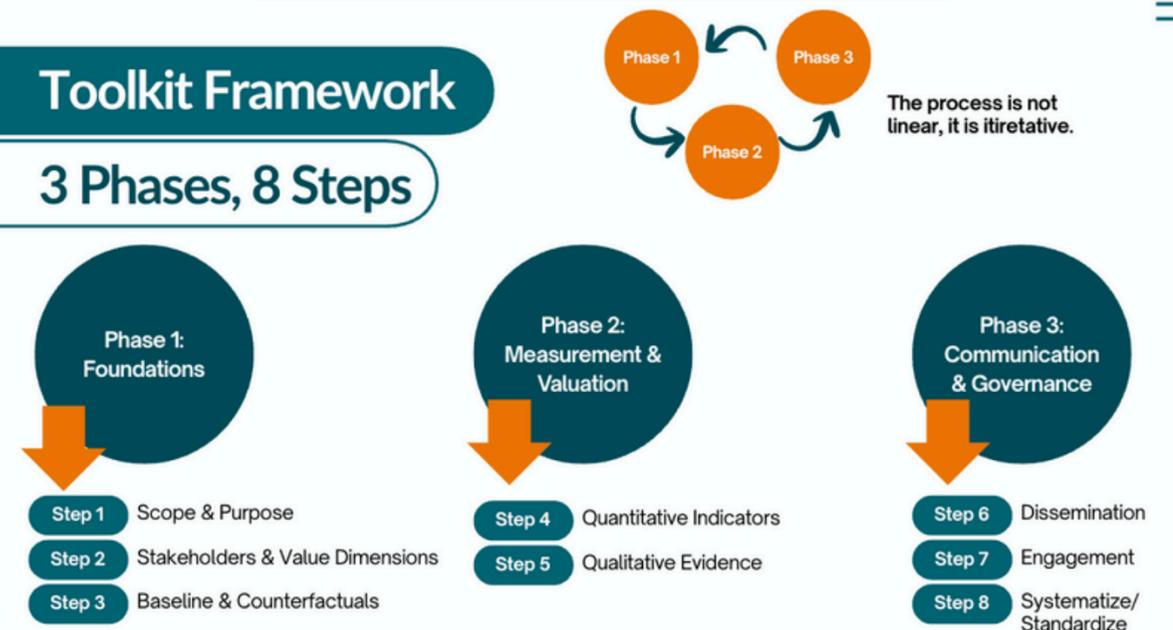
Link:

Wrapping it All Together

**In this ROI toolkit, we aim to provide a path from foundations to sustainability.**

The key here is that ROI isn't a one-time exercise; it's a practice. By following these phases and steps, we hope that you can embed ROI into your institution's culture, strengthen decision-making, and ensure long-term support for cyberinfrastructure.

Purdue University: 20 Years of ROI Storytelling

- **Early goal:** Prove cost-effective, large-scale compute capacity (2004–2006).

- **Middle years:** Demonstrate researcher adoption, faster outputs, and reduced wait times (2006-2018).

- **Later years:** Translate CI into funding, student outcomes, and sustained institutional support.

2004   **Jerry McCartney**   **2025**

Jim Bardino   **Current CIO**
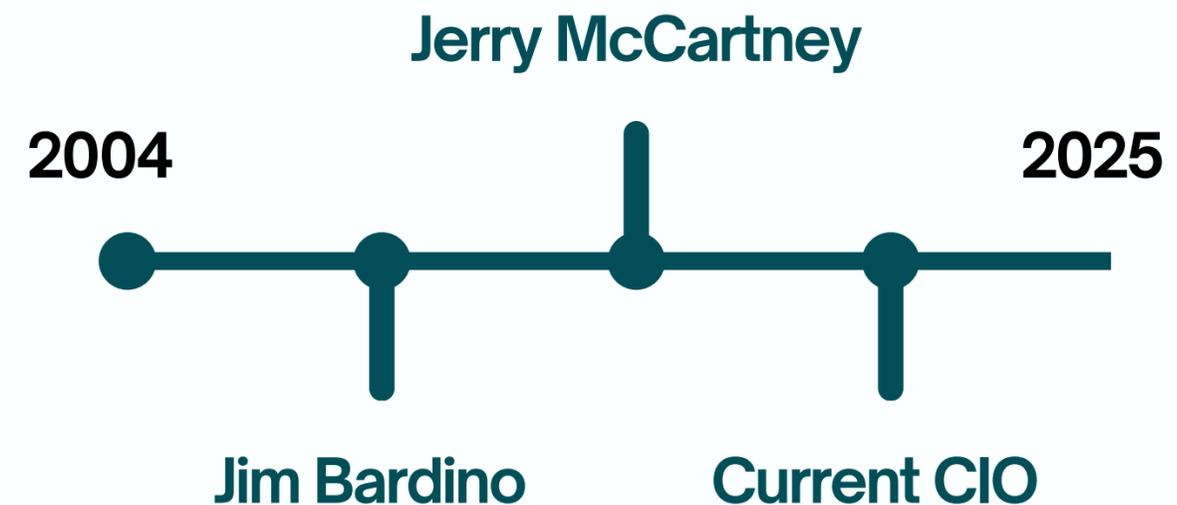
Core lesson:

The story you tell depends on who you're talking to and what's important at the institution at this point in time.

Capacity → Adoption → Strategic ROI.

Zooming In...

Early Years: 2004 - 2006

- Research computing at Purdue hadn't received much attention for years. Jim Bardino's job was to build it back up, starting with showing how little capacity existed and how expensive it was.
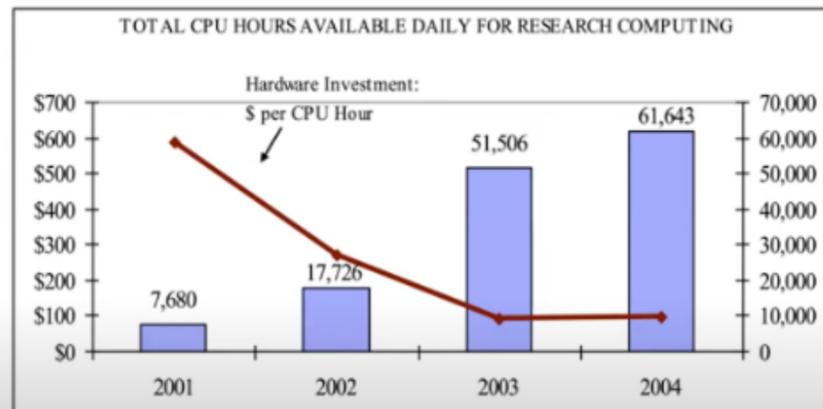
The message was clear: "If we invest wisely, we can provide more computing power for researchers while keeping costs under control."

Core lesson:

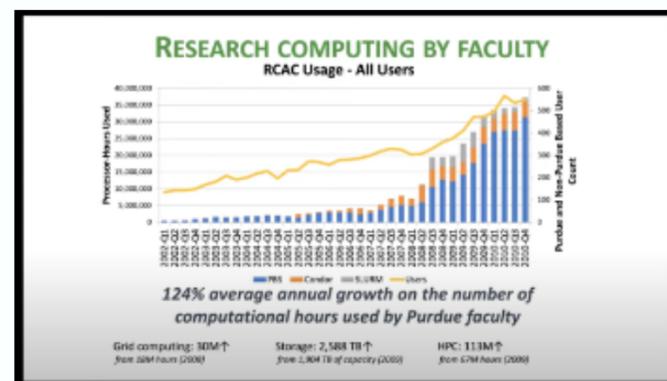**Start with baseline data and show efficiency gains**, even simple numbers are persuasive.

Mid Years: 2006 - 2018

- After investments began, they tracked who was using the systems and how much.
- Example: From 2010, the number of computational hours used by faculty grew 124% annually. They also tracked "quality of life" metrics: less wait time for jobs meant researchers could start science faster.

Core lesson:

Show adoption + impact on efficiency, usage metrics make investments tangible.

RESEARCH COMPUTING BY FACULTY
RCAC Usage - All Users

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

Grid computing: 30M↑ from 30M hours (2009)   Storage: 2,588 TB↑ from 2,904 TB of capacity (2009)   HPC: 113M↑ from 67M hours (2009)

AVERAGE WAIT TIME
ALL CLUSTERS

The message to leadership: "We're not just buying machines; people are actually using them effectively, and science is moving faster."
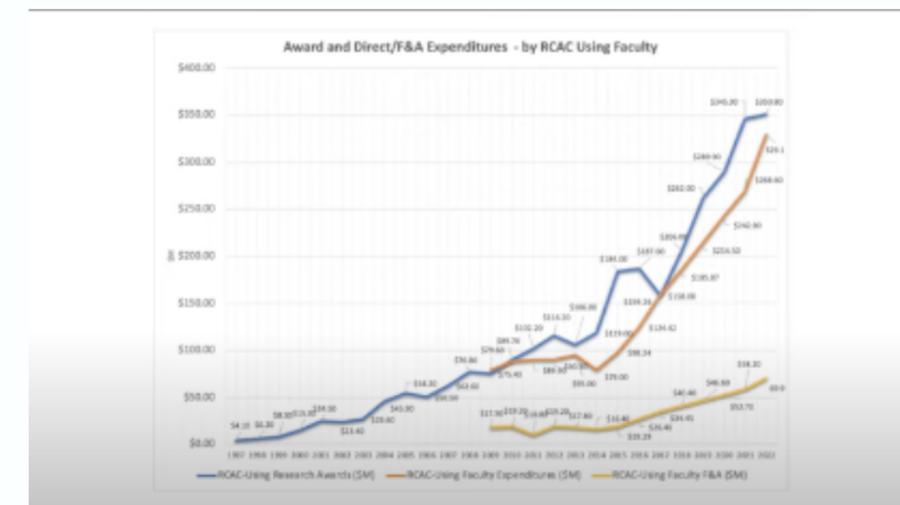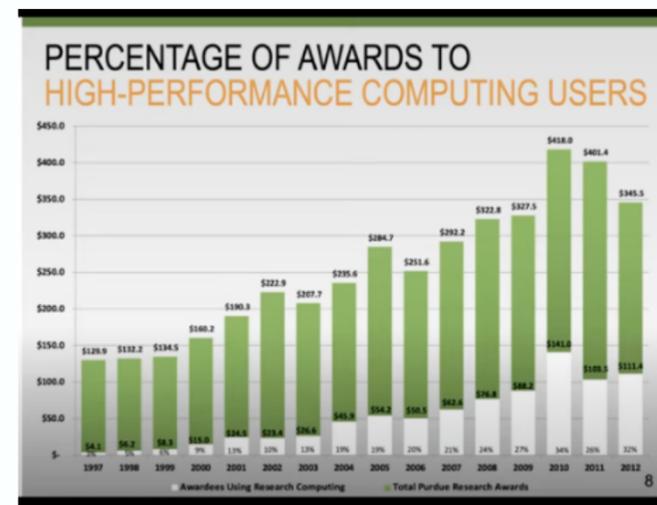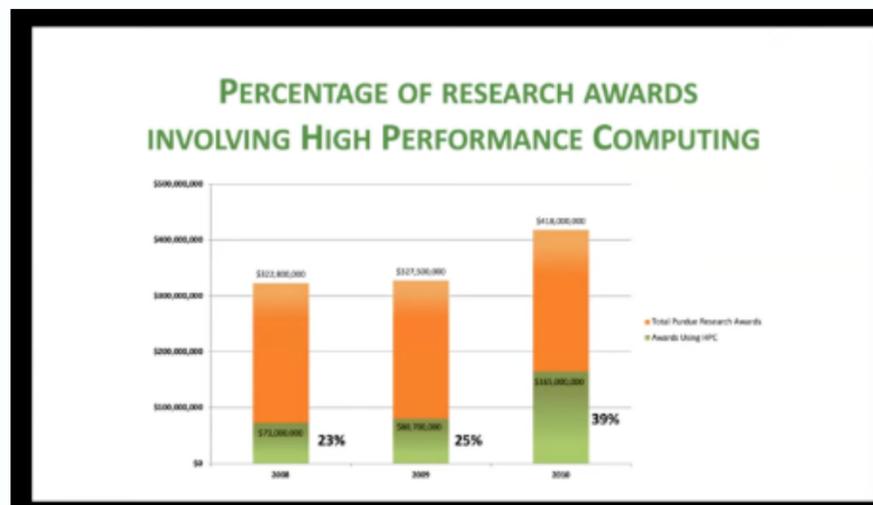
Later Years: After 2018...

Metrics over time: Financial & Research Impact
- HPC users brought in ~66% of sponsored research expenditures in 2022.
- Institutional investments like $5M in HPC led to $345M in grants (≈66× ROI).
- F&A (indirect funds) from HPC users contributed significantly to the general fund (~79% in 2022).

Core lesson:

**Show leverage**, how a small investment supports larger outcomes.

Later Years: After 2018...

Recently, Purdue added student-focused metrics:
- Tracked the number of degrees earned by HPC users by degree level.
- Example: Initially, Master's students dominated usage; now PhD students are a significant proportion.

Core lesson:

ROI is not just financial, it includes human capital and educational impact.
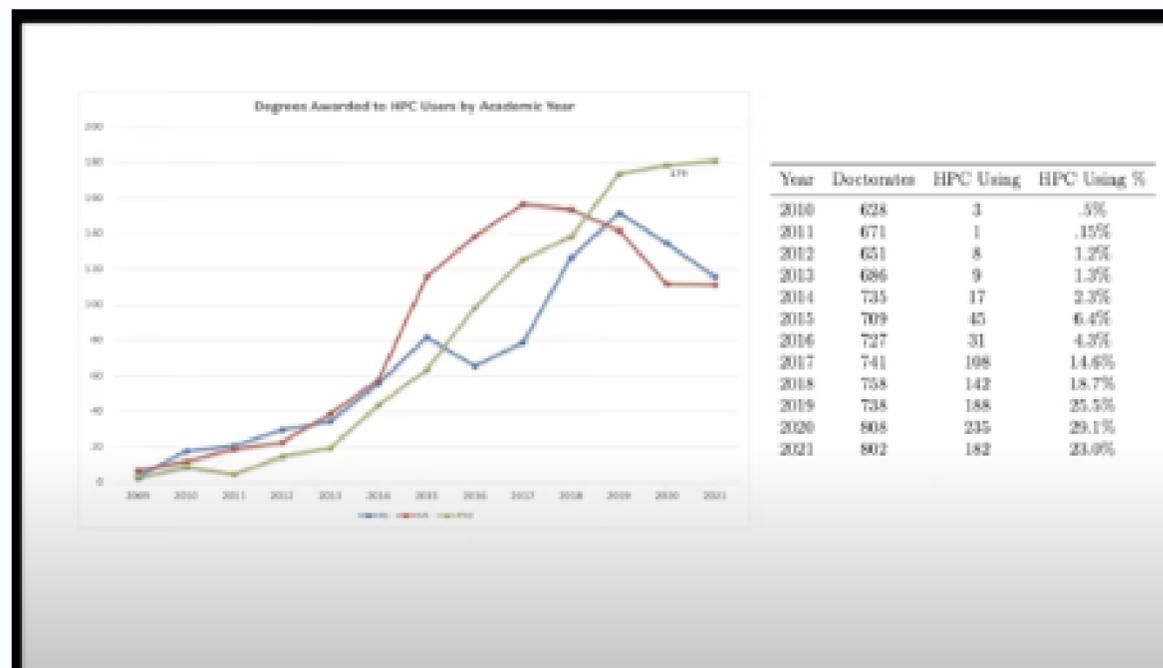
Mapping Purdue's Action to the ROI Toolkit

Purdue's ROI practice evolved over 20 years, purpose, priorities, and metrics changed over time, and they continuously built on prior work. **This demonstrates why the ROI Toolkit is iterative: frame, measure, communicate, then refine as context and stakeholders evolve.**

Phase 1: Foundations

- **Purpose & Scope**: Rebuild HPC, cut cost per CPU hour, boost research productivity
- **Stakeholders & Value**: Board/CFO (cost efficiency), Provost/President (research impact), Faculty (grants/publications), Students (training/degrees)
- **Baseline & Counterfactuals**: Low capacity, high cost → without investment: outsourcing, lost competitiveness.

Phase 2: Measurement & Valuation

- **Quantitative:** CPU hours, users, jobs, wait times, per-core cost, % awards using HPC, research expenditures, degrees, F&A contributions
- **Qualitative**: Faster publications, collaborations, student outcomes, improved researcher experience

Phase 3: Communication & Governance

- **Disseminate:** CFO (financial leverage), Board (top researcher adoption, faculty usage impact), Provost/President (student success)
- **Engage:** Dashboards, briefings, metrics evolve with priorities
- **Institutionalize:** Shared ownership, standardized metrics, link to budget, long-term tracking.

References &

Other Helpful Resources

- Snapp-Childs, W. G., Hart, D. L., Costa, C. M., Wernert, J. A., Jankowski, H. E., et al. (2024). Evaluating return on investment for cyberinfrastructure using the International Integrated Reporting framework. *SN Computer Science*, doi:https://doi.org/10.1007/s42979-024-02889-z
- Smith, J. (2022, September 21). Metrics for RCD success [Slides]. CaRCC Emerging Centers Call. https://www.youtube.com/watch?v=L7dTYQooe8A&list=PLV-SC0CHLTwcAP6x9NbHeqaVNSYshCY_H&index=22&t=1750s
- Smith, P. (2024) The value proposition of campus high performance computing facilities to institutional productivity: A production function model. *SN Computer Science. 5*, 551. https://doi.org/10.1007/s42979-024-02888-0
- Stewart, C., Wernert, J., Costa, C., Hancock, D. Y., Knepper, R., & Snapp-Childs, W. (2022, July). Return on investment in research cyberinfrastructure: State of the art. In Practice and Experience in Advanced Research Computing (PEARC '22). Association for Computing Machinery. https://doi.org/10.1145/3491418.3535131

Midwest RCD Consortium

THANK YOU

Esen Gokpinar Shelton, Ph.D.[1]; Winona Snapp-Childs, Ph.D.[1]; Preston Smith, Ph.D.[2]

[1]Indiana University, [2]Purdue University

The toolkit may not be copied, distributed, or used without explicit permission from the authors.

Contact: pti@iu.edu

https://midwestresearchcomputing.org/