

Shared CI Services: Opportunities and Challenges

Moderator: Adam S. Bolton
NSF's National Optical Astronomy Observatory





Pamela Hill, UCAR / National Center for Atmospheric Research



JJ Kavelaars, NRC Herzberg, Canadian Astronomy Data Centre



Von Welch, IU, Center for Applied Cybersecurity Research



Mike Zentner, SDSC, Science Gateways Community Institute + HUBzero

What are the opportunities and challenges for sharing of the following services across facilities?

- Identity Management
- Storage-as-a-Service
- Disaster Recovery
- Computing-as-a-Service
- Data Discovery
- Others



Domain-specific v. shared CI

Astronomy
Biology
Chemistry
Earth Science
Physics
Etc.

Compute
Storage
Auth
Data Discovery
Etc.



Domain-specific v. shared CI

Astronomy
Biology
Chemistry
Earth Science
Physics
Etc.

Compute
Storage
Auth
Data Discovery
Etc.



Domain-specific v. shared CI

**Software
Systems**

**Cyber-
Infrastructure**

Software on CI

Software as CI



What is the sharing model?

- Organizational
 - E.g., ITIL/COBIT model for shared IT service & governance
- Mission-driven
 - E.g., like DOE-funded CI centers
- Partnership-based
 - E.g., partnership agreements with major NSF-funded CI centers
- Vendor-oriented
 - E.g., AWS, Azure, Google Cloud
- Collaborative
 - E.g., facilities all run and share CI with one another?



Test case: NSF's National Center for Optical and Infrared Astronomy



What are the opportunities and challenges for sharing of the following services across facilities?

- Identity Management
- Storage-as-a-Service
- Disaster Recovery
- Computing-as-a-Service
- Data Discovery
- Others



Shared CI Services Opportunities and Challenges

NSF CI Workshop – September 2019

Pamela Hill

Manager HPC Data Infrastructure Group



UCAR / NCAR

- University Corporation for Atmospheric Research (UCAR)
 - Our 117 member colleges and universities set UCAR's overall direction, ensuring that we are meeting the most urgent needs of the research and education community. The members, who represent virtually all the major Earth system science education programs in North America, guide UCAR's course by electing the Board of Trustees and serving on a number of governance and advisory committees.
- National Center for Atmospheric Research (NCAR)
 - NCAR was established by the National Science Foundation in 1960 to provide the university community with world-class facilities and services that were beyond the reach of any individual institution.
 - More than a half-century later, we are still delivering on that mission. NCAR provides the atmospheric and related Earth system science community with state-of-the-art resources, including supercomputers, research aircraft, sophisticated computer models, and extensive data sets.
 - NCAR also provides rich education and outreach opportunities, from fellowships for early career scientists to free public lectures to scientific workshops.

Data Centric Design Concept

- High-bandwidth network infrastructure
 - Both within facility and to external facilities
- Minimize data movement
 - Data is available to all resources from a central point
- Support entire data life cycle
 - Computational data production
 - Data pre/post-processing
 - Data analysis
 - Visualization
 - Publication
 - Long-Term Preservation

Computational Environment

- Cheyenne – Large Computational Resource
 - 4032 nodes, EDR IB
- Casper – Smaller Data Analysis, Data Management, Visualization Resource
 - 32 nodes, GPGPU, 100Gb Ethernet, HDR IB
- Glade – Centralized Storage Resource
 - 40GbE, EDR IB
 - 38 PB capacity, 25 PB data today
- Campaign Store – Warm Archive (2018)
 - 40 Gb Ethernet
 - 26 PB capacity, 11 PB data today, 17 PB coming
- HPSS – Deep Archive, Tape
 - 100 PB data today

Data Services

- Globus infrastructure for data transfer and basic data management
- GUFI for data management
- CMIP Analysis Platform
 - Data analysis of CMIP data sets
- Science Gateways for data distribution and data discovery
 - Digital Asset Services Hub (DASH)
 - Research Data Archive (RDA)
 - Earth System Grid (ESG-F)
 - Climate Data Gateway (CDG)
- Disaster Recovery Services
 - Critical data collections
 - Cooperative agreement with University of Colorado
 - Exploring cloud resources for DR

Shared Infrastructure Challenges

- Network connectivity is critical for shared services
- Data movement of large data sets
- Data formats may not be compatible with storage resources
- Ensuring that security policies are honored (authentication vs authorization)
- Compatibility of authentication systems
- Establishing trust relationships between facilities
- Maintaining administrative boundaries (accountability)
- Predictability of model run times and job turn around times
- Ability to support complex model configurations
 - Multiple job steps, in-situ analysis, data assimilation, ML components

Canadian Astronomy Data Centre

Thoughts on Shared Services

Sept-2019 / JJ Kavelaars



National Research
Council Canada

Conseil national de
recherches Canada

Canada

CADC – Shared Service for Astronomy



- › Grew organically from hosting 1 data collection to now hosting dozens.
 - Currently host data from over 200 different instruments with staff of 23 - 7 research astronomers, 1 project manager, 4 operations staff and 10 developers and a PI
- › Expanded collections required focusing on the core services we knew how to deliver.
- › Built out our knowledge of astronomy and maintain a very tight connection to our science community (7 research astronomers using the archive to do research)
- › Each member of staff works to service the individual needs of our users.
- › Services are expanded when user demands requires.

CANFAR and IVOA

- › CANFAR infrastructure built with IVOA standards



- › Early implementers/adopters to verify standards meet needs
- › Advocating support for authorisation in all services

ADQL	SSO
CDP	TAP
DALI	TAPRegExt
DataLink	UWS
ObsCore	VODataService
RegistryInterfaces	VOResource
SIA	VOSI
SimpleDALRegExt	VOSpace
SODA	VOTable



CANFAR in 2018

Number of files ingested	10,357,541
TiB ingested	120
Number of retrievals	47,559,651
TiB retrieved	737
Jobs submitted	6,118,009
Core years used	1,154
DOIs issued	16
Authenticated users retrieving	236
Number of IPs with retrievals	15,446

Number of user groups	478
OpenStack users	186
OpenStack projects	53
Batch users	75
Peak retrievals/day	2,457,618.0
Peak TiB retrievals/day	60
Peak file ingestion/day	378,074.0
Peak TiB ingestion/day	9

Multiple Levels of Shared Services in actions



- › NRC Canada – 3700 staff across 10-15 Institutes
 - Building Shared Services - APSM
 - Manufacturing Shared Services - DFS
 - IT Share Services – KITS
 - Human Resources Shared Service – Common Services
- › Many different layers within the organization that cut across different research programs - Previously each program maintained their own groups in these areas. Consolidation removed connections between these service staff levels and the research goals and strategic directions of the institutions.
- › Services staff at local sites under pressure to follow centralized management.
- › Local staff have lost access to important expertise.

CADC – Shared Services Canada



- › Shared Services Canada – Created by the stroke of a pen and consolidated the activities of hundreds of departmental systems.
- › No plan for how to manage this transition was developed.
- › No clear authorities for who could make what decisions.
- › Groups in regions were immediately absorbed into centralized structure – loyal staff attempted to ‘work around’ the structure to ensure critical operations remained in place.
- › Key goal : Consolidate 800 data centres into 4, serve 500,000 email boxes from one sites, merge all telecommunications into a single centrally managed system.
- › This was an act-now plan later initiative.

CADC – Disaster Recovery



- › We have a plan for recovering the CADC from a disaster using our internal staffing and expert knowledge.
 - Relies on ‘near-line’ storage infrastructure housed within Compute Canada (one of our Common CI provider)
- › “Are there services such as disaster recovery, identity management, computation, and others that can be shared across facilities?

CADC - Identity Management



- › Currently have our own internal Identity Management system.
 - Have examined moving to eduGain but our user community extends outside the academic one.
 - Desired to provide strict compliance with Global Data Privacy Rights makes us reluctant to use commercially provided identities.
 - Our identities provide access to resources that we control. Desire to have flexibility regarding who we choose to provide those resources to.

CADC – Computation Service

- 100% of our research computation is provided by Compute Canada as a shared service.
 - CC is a partner who interacts and dialogs with CADC on our community needs.
 - Work together (CADC directly provides funding to CC to enable them to support our requirements)
 - Initial attempt at partnership took 2 years to develop and resulted in cancelled contract with upper levels of organization.
 - Visibility into processes via open sharing of planning tools (Jira and GitHub)

CADC – Archive Service



- › Archive systems are provided by and maintained by Shared Services Canada.
 - Developing a partnership with SSC to 5 years! Attempts to partner across the top of the organization failed.
 - Eventually pushed our connections down into the organization to establish partnership.
- › Cost of current system about double what we know we could do for ourselves.
- › Staff are managed externally so communication can be in-efficient.
- › Low visibility into the processes or activities of our SSC partners.

NRC Herzberg – Cautionary advice



- › Make sure your shared-services are provided by a peer level partner who understands your business case and wants to help solve that business case.
- › Going to a shared infrastructure has not, for us, reduced costs.
- › SSC and CC now, after much effort, providing us with access to expertise and experience that we would not otherwise have.

NRC – Herzberg Share Service problems



- › Lack of local ‘on the ground’ staff is problematic.
- › Lack of awareness of operational realities –
 - Windows vs. Linux,
 - Research Network vs. Secured Government Network
- › Lack of understanding the business case as been very frustrating.
 - Clients not other parts of Government!
 - Clients don’t use closed systems.
 - Data is free
- › Solutions not matching requirements (we are small part of requirements)
 - Cell phones instead of offices phones for a radio telescope facility!
 - Closed VC systems
 - Root access to the desktop! [no VMs allowed!]



I Cut, You Choose: How to Share a Service?

2019 NSF Large Facilities CI Workshop
Shared CI Services: Opportunities and Challenges Panel
Von Welch, vwelch@iu.edu



**CENTER FOR APPLIED
CYBERSECURITY RESEARCH**
PERVERSIVE TECHNOLOGY INSTITUTE



ResearchSOC

- Operational services and related training for NSF CI
- Community of Practice and Threat Intelligence Network
- Enabling Cybersecurity Research
- Outreach to Higher Ed Infosec regarding research CI



- Creating comprehensive cybersecurity programs
- Community building and leadership
- Training and best practices
- Tackling specific challenges of cybersecurity, software assurance, privacy, etc.

Trusted CI: Impacts

Trusted CI has positively impacted over 260 NSF projects since inception in 2012.

Members of more than 180 NSF projects have attended our NSF Cybersecurity Summit.

Members of more than 80 NSF projects have attended our monthly webinars.

We have provided more than 300 hours of training to the community.

We've had engagements with 41 projects, including nine NSF Large Facilities.

Convene Large Facilities Security Team with all 23 LFs represented.



The Trusted CI Broader Impacts Project Report

June 28, 2018

For Public Distribution

Jeannette Dopheide¹, John Zage², Jim Basney³

<https://hdl.handle.net/2022/22148>

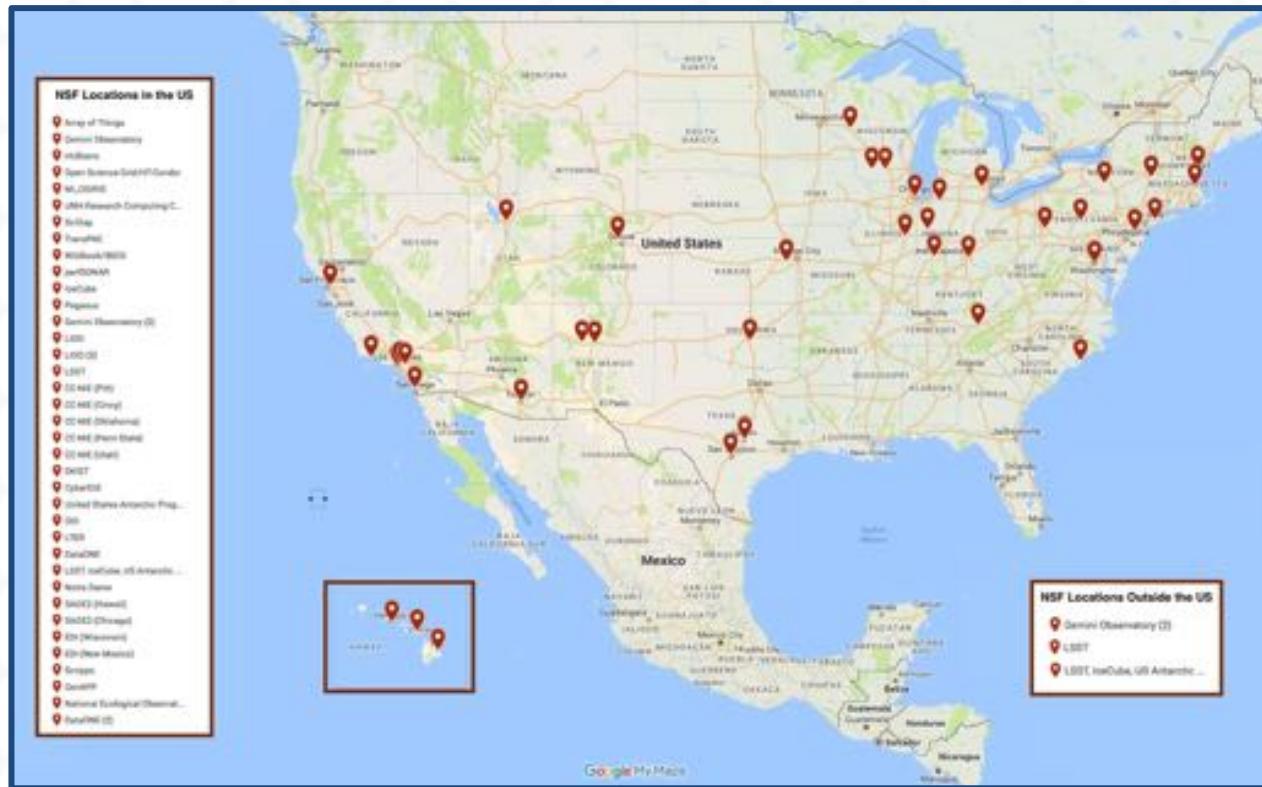
Engagements: One-on-one Collaborations

We take applications every six months.

Accept applications every six months:

<https://trustedci.org/application/>

Deadline: October 2, 2019



Annual NSF Cybersecurity Summit

One day of training and workshops.

Lessons learned and success from community.

Oct 15-17, 2019 in San Diego

<https://trustedci.org/summit/>

Agenda driven by call for participation, due August 12th

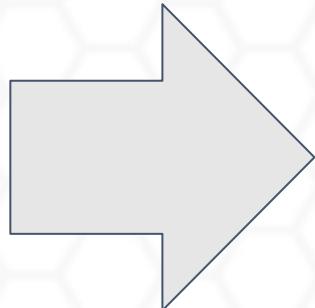
<https://trustedci.org/cfp2019>



ResearchSOC: Operational Cybersecurity Services for Science



STINGAR

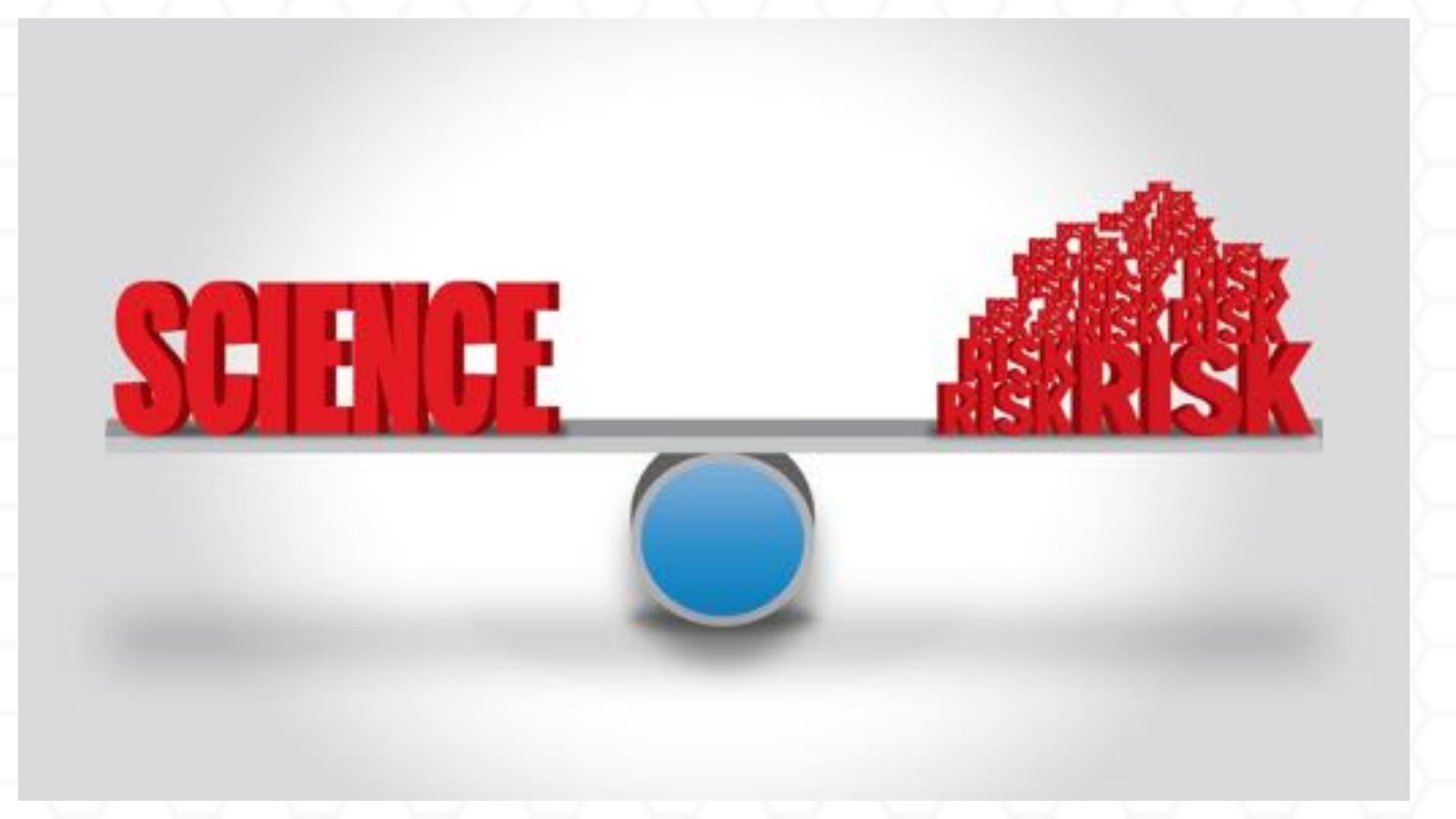


GAGE
UNAVCO



NATIONAL RESOURCE FOR TRANSLATIONAL
AND DEVELOPMENTAL PROTEOMICS

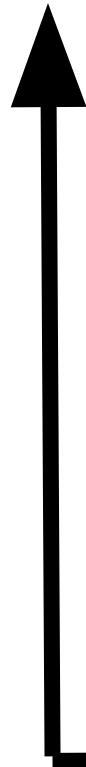
“Are there services such as disaster recovery, identity management, computation, and others that can be shared across facilities?”



SCIENCE

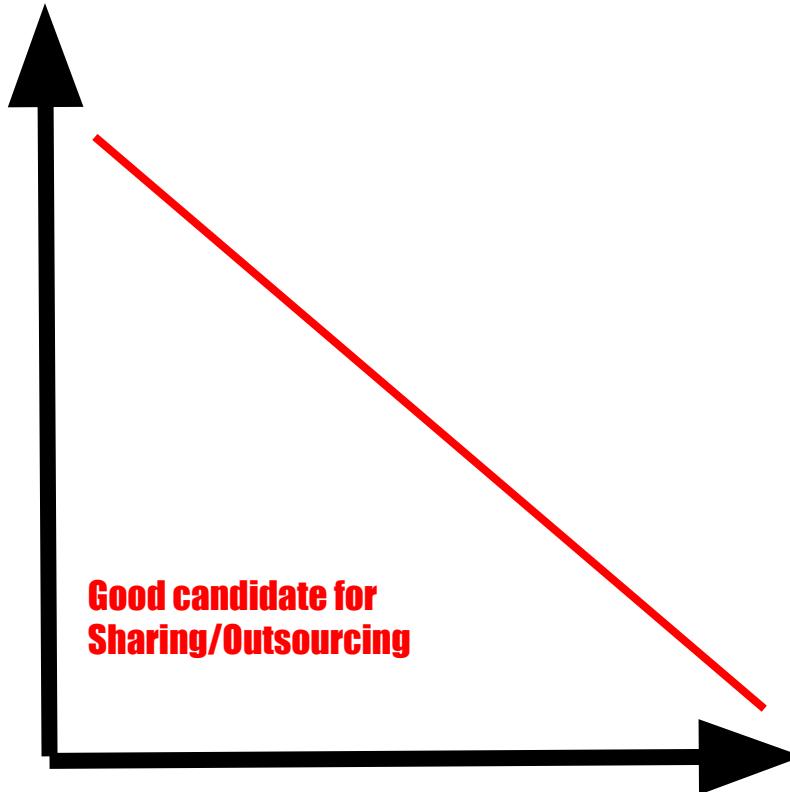
RISK

Specific



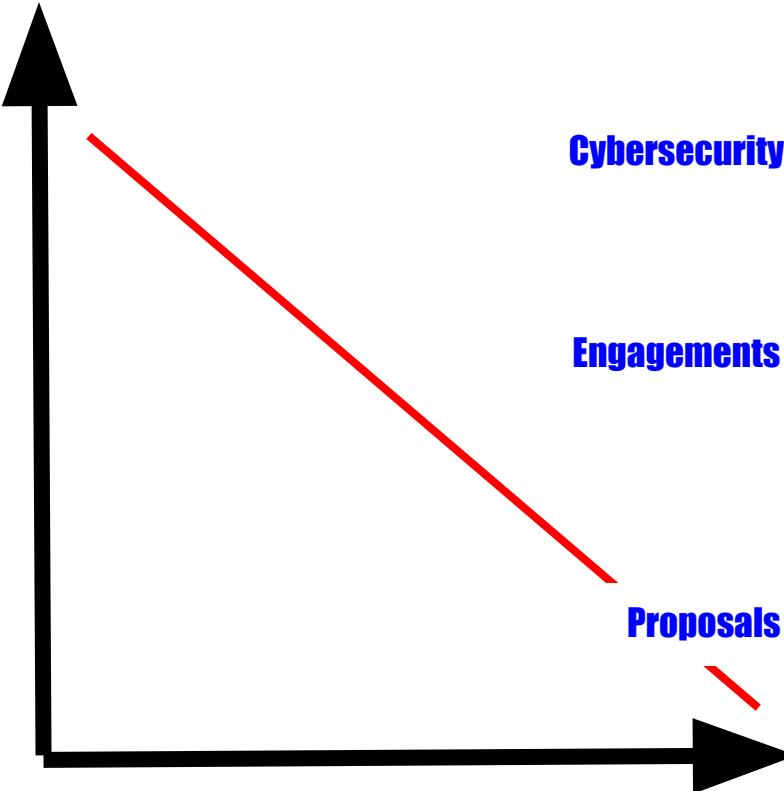
Critical

Specific



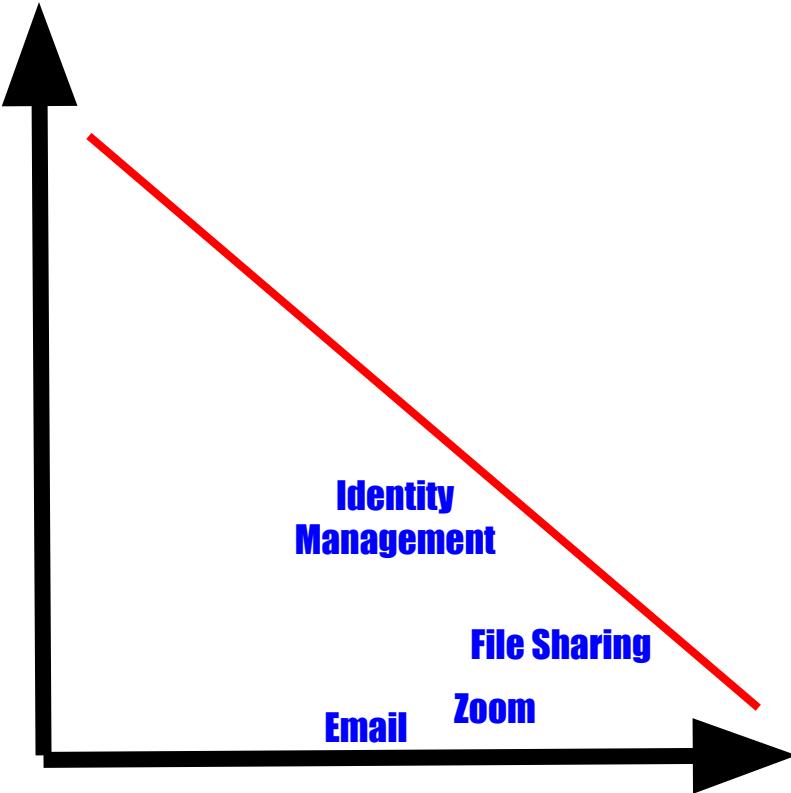
Critical

Specific



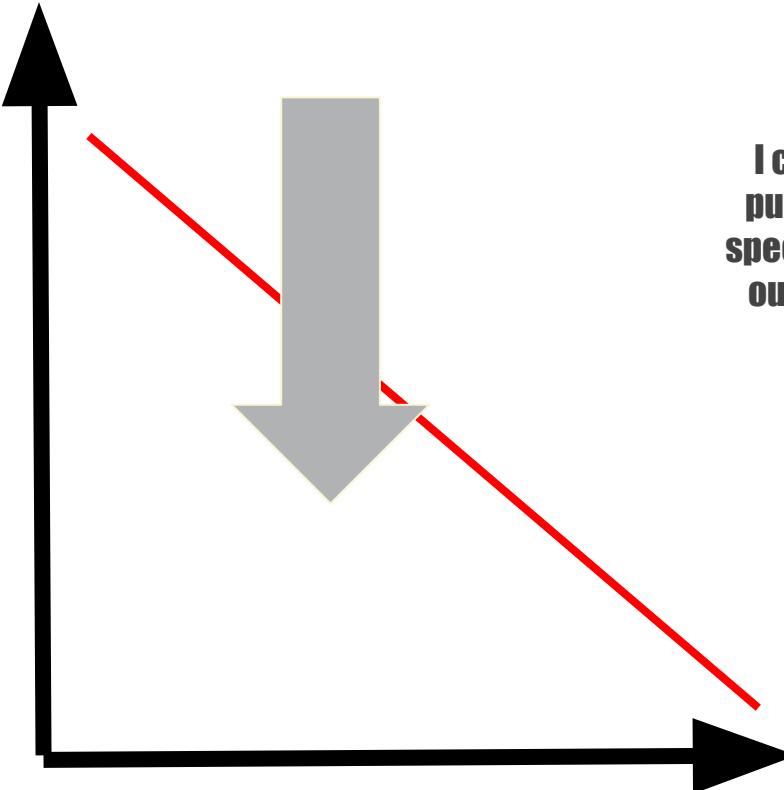
Critical

Specific



Critical

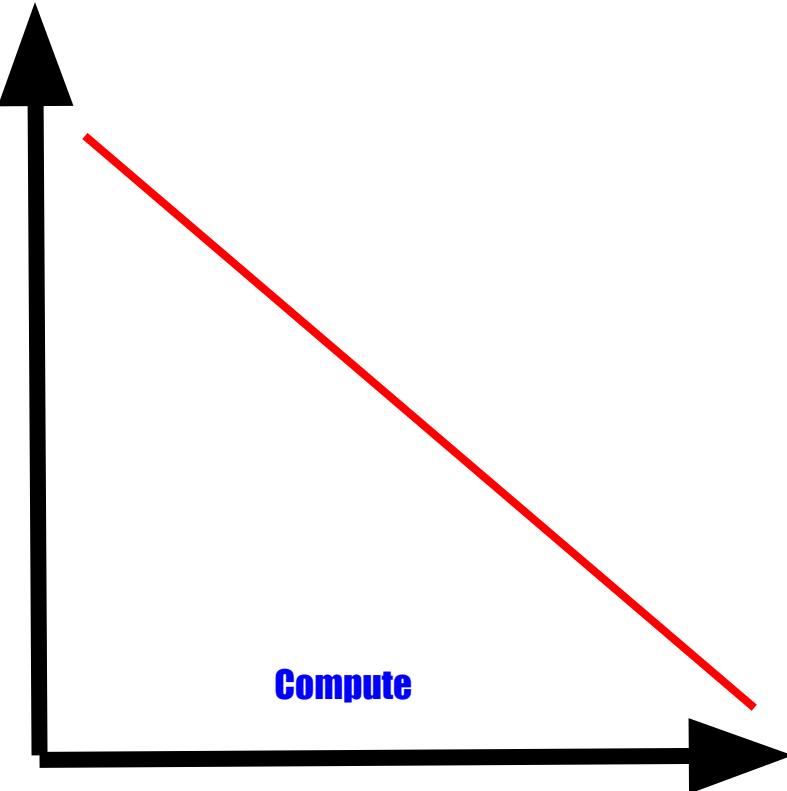
Specific



**I constantly
push against
special so I can
outsource it.**

Critical

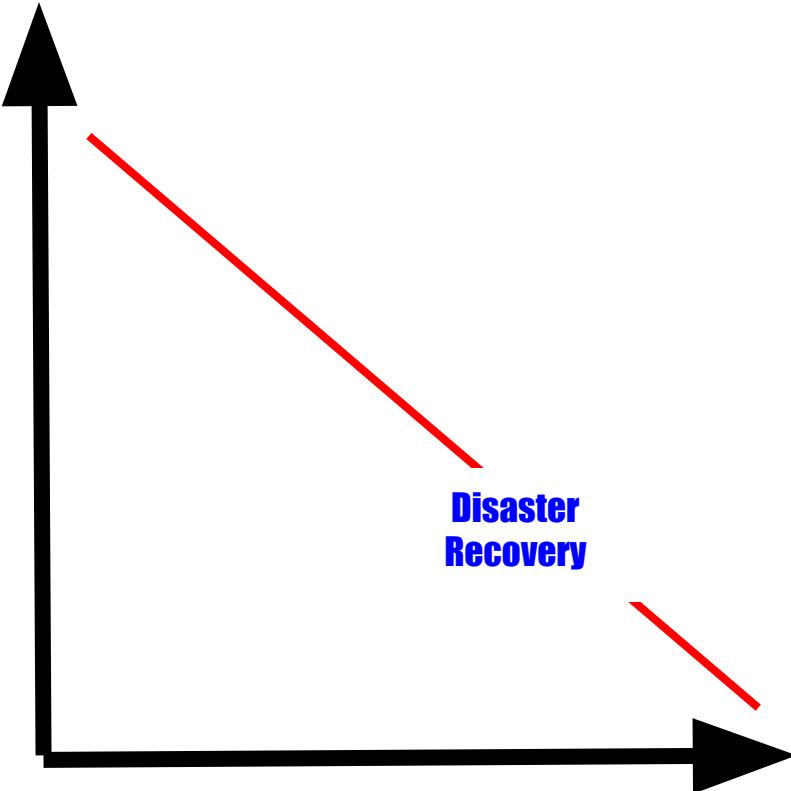
Specific



**Standard
workflows and
architectures
(HTC, HPC, IaaS,
GPUs, etc.)**

Critical

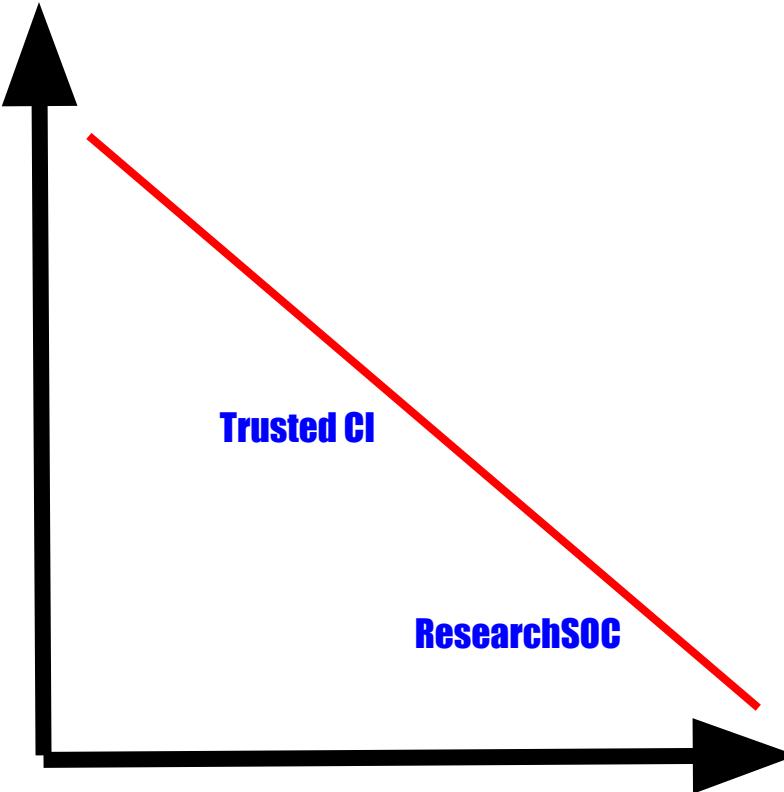
Specific



Very
preliminary
thinking here...

Critical

Specific



Critical

The Other Key Factor:
Who Runs a Service is as Important as What It Does

Interested in Supporting Research
Rather Than Doing Research

Capable of Weaving Together
Sustained Funding

Capable of Weaving Together
a Sustained Workforce

Servant and Leader:
Do Want I Want When I'm Right.
Tell me that I'm Wrong when I'm Wrong.

Thank you.

Trusted CI is supported by the National Science Foundation under Grant 1547272. ResearchSOC is supported by the National Science Foundation under Grant 1840034. The views expressed do not necessarily reflect the views of the National Science Foundation or any other organization.



Connecting people and resources
to accelerate discovery by empowering
the science gateway community

A large, abstract network graph is positioned in the upper right corner of the slide. It features numerous small, semi-transparent circular nodes in shades of purple, blue, and pink, connected by a web of thin, light-colored lines. The nodes are more densely packed towards the center and become more sparse towards the edges.

Large Facilities CI Workshop Shared CI Services: Opportunities and Challenges

*Michael Zentner, San Diego Supercomputer Center
Director, Science Gateways Community Institute
Director, Sustainable Scientific Software
Director, HUBzero Platform*

sciencegateways.org

Award Number
ACI-1547611



A close-up photograph of a person's hands holding a red apple. The apple is cut in half, revealing its bright yellow-green flesh and a large, prominent central seed. The hands are positioned to showcase the fruit, with fingers visible around it. The background is blurred, showing hints of other colors and shapes, possibly suggesting a market or a workshop environment.

Plenty of
opportunities to
share not just
technologies but
also competencies



TECHNICAL

- Domain Scientists
- HPC
- Middleware
- Security
- UX Design
- Web Design
- Web Programming
- Analytics
- Databases
- Microservices
- Container Orchestration

What does one need to run CI?



TECHNICAL

- Domain Scientists
- HPC
- Middleware
- Security
- UX Design
- Web Design
- Web Programming
- Analytics
- Databases
- Microservices
- Container Orchestration

OPERATING

- System Administration
- Devops
- Hardware
- Cloud Configuration
- Testing
- Continuous Integration
- Version Management
- Disaster Recovery

What does one need to run CI?



What does one need to run CI?

- Domain Scientists
- HPC
- Middleware
- Security
- UX Design
- Web Design
- Web Programming
- Analytics
- Databases
- Microservices
- Container Orchestration

OPERATING

- System Administration
- Devops
- Hardware
- Cloud Configuration
- Testing
- Continuous Integration
- Version Management
- Disaster Recovery

- Project Management
- Outreach
- Marketing
- Workforce Development
- Sustainability Planning
- Revenue Collection
- Education
- Impact Measurement
- Evaluation



TECHNICAL

- Domain Scientists
- HPC
- Middleware
- Security
- UX Design
- Web Design
- Web Programming
- Analytics
- Databases
- Microservices
- Container Orchestration

OPERATING

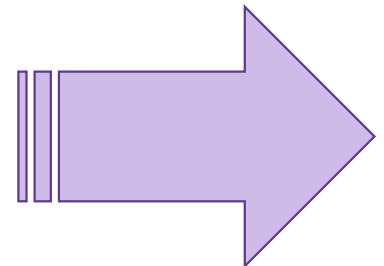
- System Administration
- Devops
- Hardware
- Cloud Configuration
- Testing
- Continuous Integration
- Version Management
- Disaster Recovery

- BUSINESS
- Project Management
 - Outreach
 - Marketing
 - Workforce Development
 - Sustainability Planning
 - Revenue Collection
 - Education
 - Impact Measurement
 - Evaluation



Why should we consider sharing competencies?

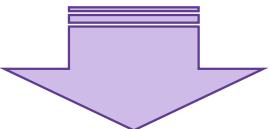
- Efficiency
- Expertise
- Connections
- (other?)



- Lower Costs / Volume Discounts
- Focus on core competencies
- Grow Audience Faster and Larger
- Novel Intellectual Property



Getting to Market Faster



Being More Sustainable

The most sustainable projects might be those that develop no software



**The most
sustainable
projects might
be those that
develop no
software**



Nevertheless, question(s) we should answer: Why don't we share competencies (more)?

- I can do it better if I redevelop...
...that thing that people have spent 10 years learning the hard lessons already.
- I would be admitting inferiority?
- Control, lack of ability to collaborate?
- Religious technology debates?
- Stigma of spending money elsewhere?
- Other?



What makes us ignore these questions?

Day 1



Day 1





Day T-1



Where I should not invest

Where I should invest

Day 1



Day N



Example:
SGCI



Science Gateway Pain Points

Connecting expensive resources & data to

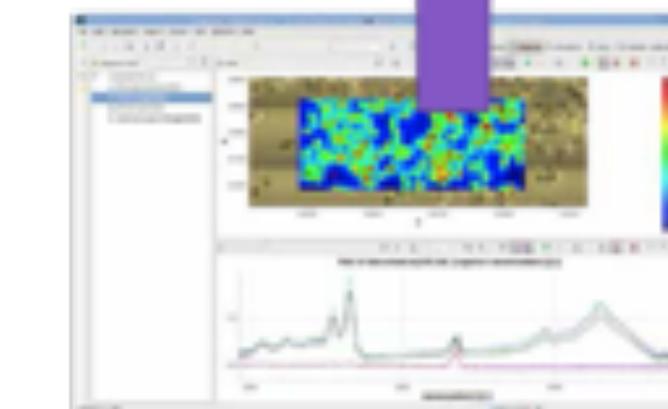
scientist authors that write codes to utilize them to

audiences that use those codes, data, and instruments en masse

(expensive) Resources



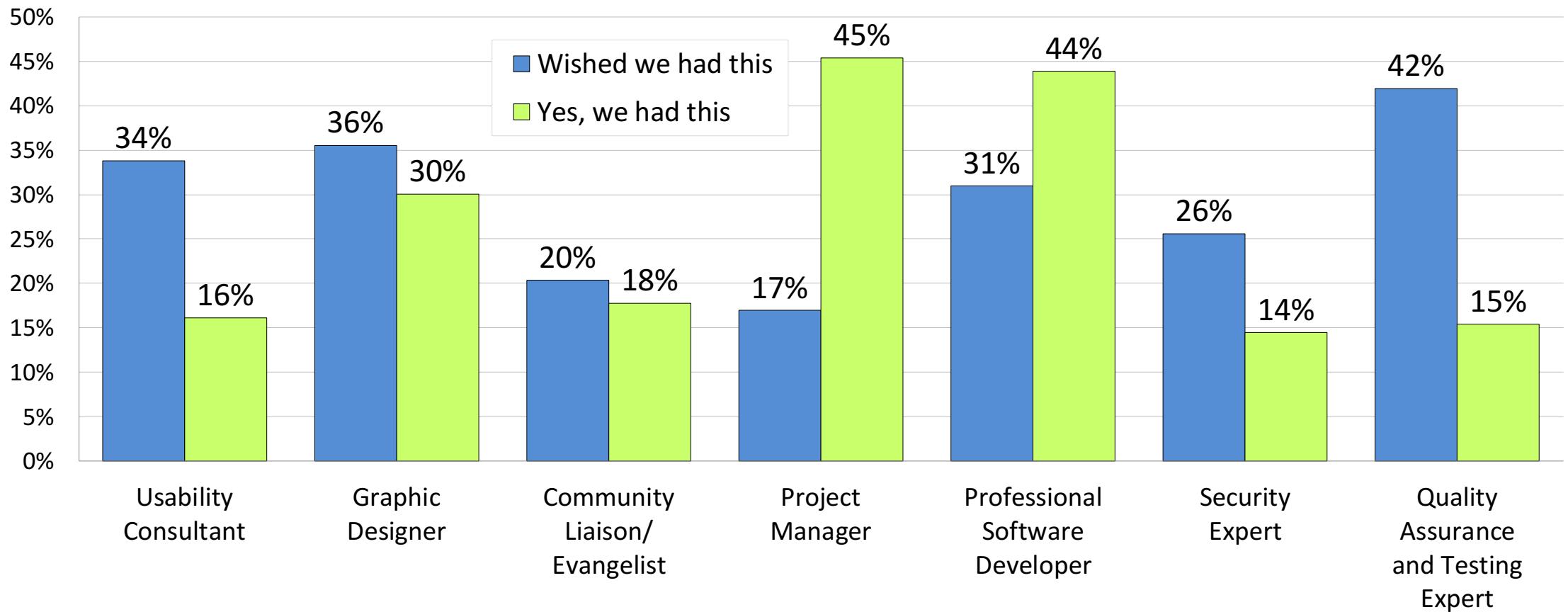
Audience(s)



Collaborating Scientist Software / Data Authors



We learned that diverse expertise is needed





The formation of
an institute...
to institutionalize
best practices

Community
Knowledge
Awareness
Efficiency
Sustainability



Science Gateways Community Institute

Products
Thought
Leadership



- ✓ **Technology consulting**
 - ✓ Developer placement
 - ✓ Technology selection
 - ✓ Cybersecurity (w/TrustedCI)
 - ✓ Usability & graphic design
- ✓ **Operations & business consulting**
 - ✓ Marketing approach development
 - ✓ Licensing consulting
 - ✓ Activity based budget & resource planning
 - ✓ Funding / sustainability planning
- ✓ **Education & community**
 - ✓ Gateway “Focus Week”
 - ✓ Annual Gateways conference
 - ✓ Gateway catalog
 - ✓ Hackathons
 - ✓ Internships
 - ✓ Gateway & science ambassadors
 - ✓ Technical forums

The numbers show interest in shared services

771

webinar
attendees



1509

participants
in SGCI's
events

210
Bootcamp
attendees

13
Affiliates

32
success
stories



32
Consultations

283
student &
41
faculty
participants

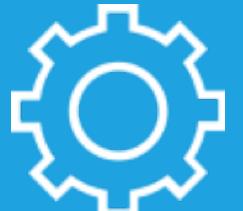


106
letters of
collaboration

13
partners

35

Extended
Developer
Support
projects



\$613,845
additional funding to SGCI
by external projects

586
Gateway Catalog entries





Also promising from a sustainability standpoint

Recognized Value of Services

\$614,000

Additional Funding from External Projects

\$25,000

Additional Funding from Focus Week Events

Demonstrated Value Propositions

41%

Collaborating Project Hit Rate

8.5x

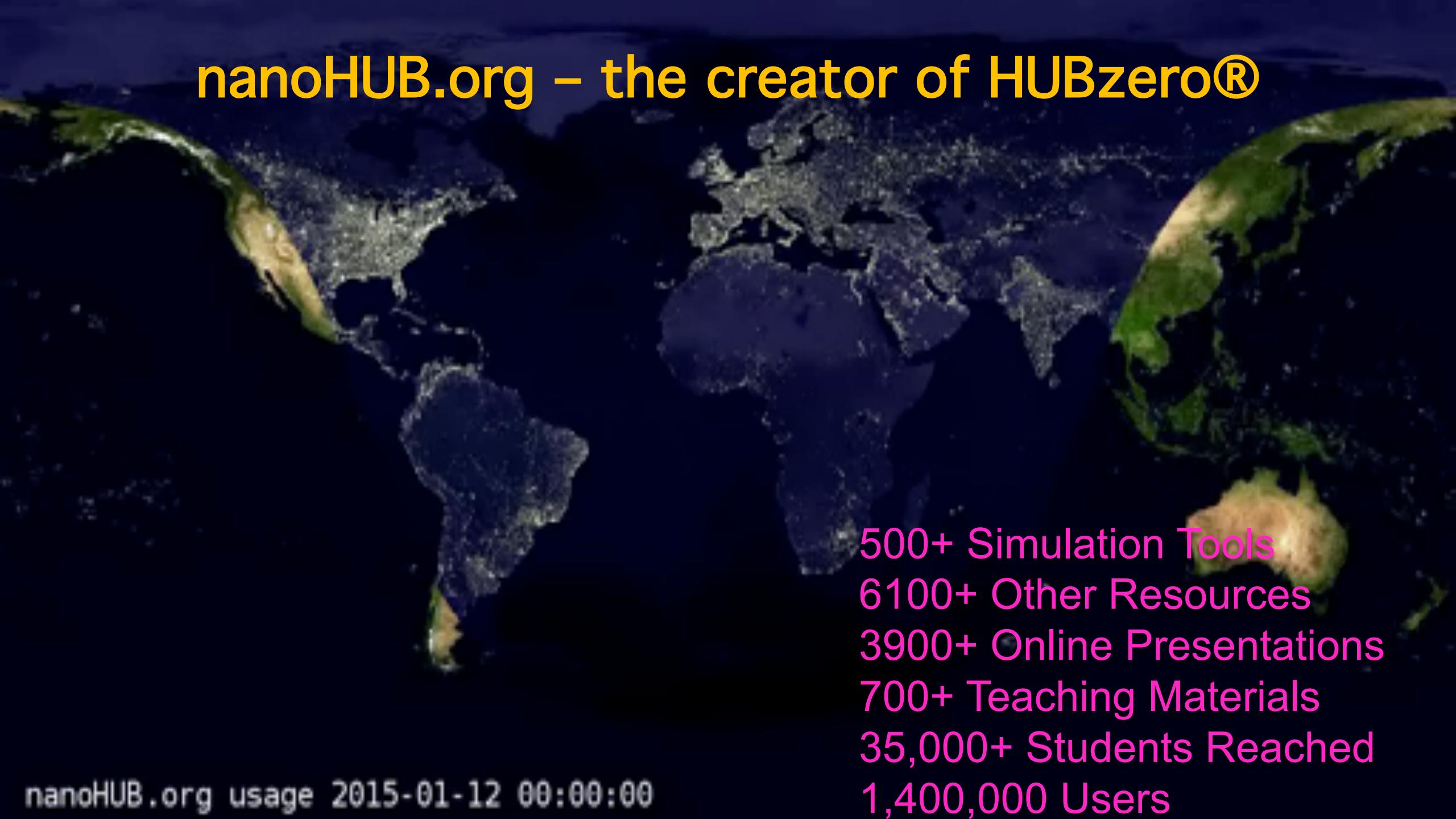
Demonstrated Return on Investment
More Money for Science

Nearly All Services Have A Waiting List

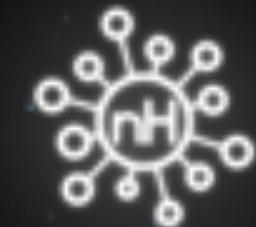
A close-up photograph of a person's hands holding a red apple. One hand is gripping the stem end, while the other holds a wedge-shaped slice. The apple has a vibrant red skin with some yellow and green at the core area. The background is blurred.

Example:
HUBzero®

nanoHUB.org – the creator of HUBzero®

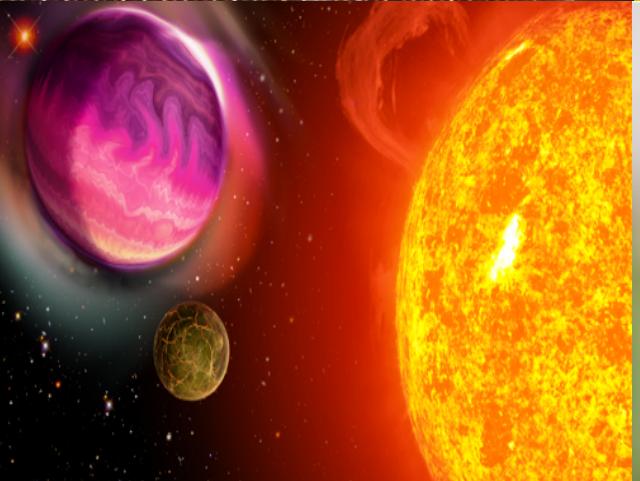
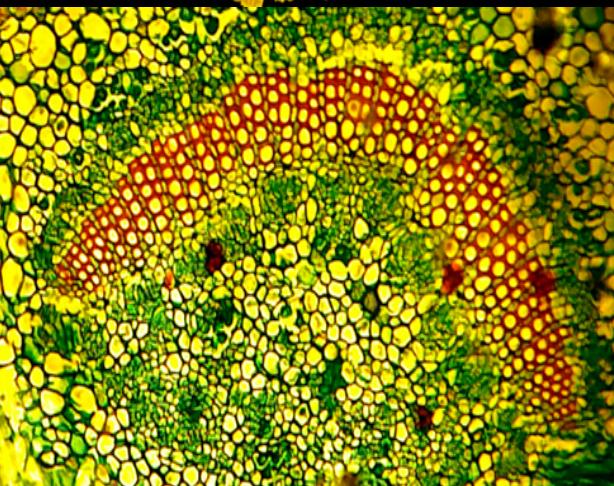
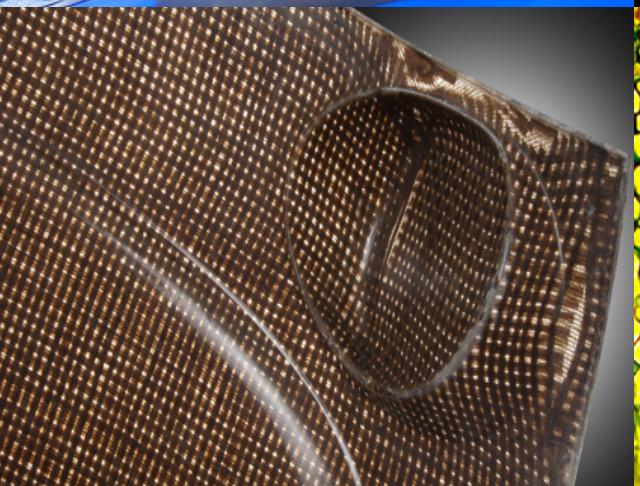
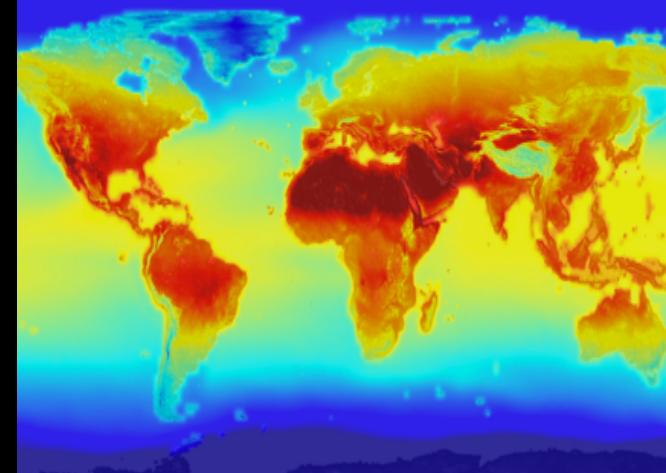
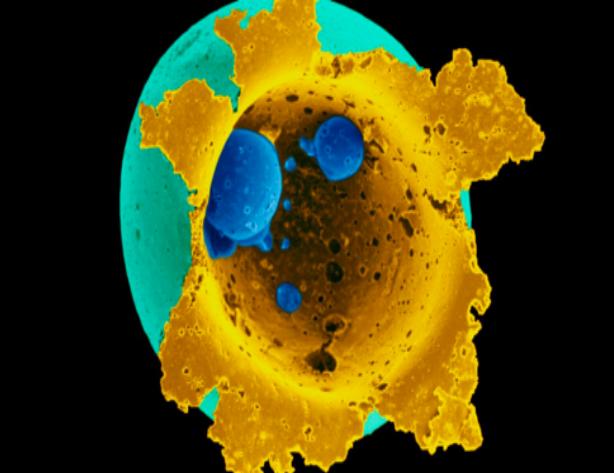


500+ Simulation Tools
6100+ Other Resources
3900+ Online Presentations
700+ Teaching Materials
35,000+ Students Reached
1,400,000 Users



Spun off in 2007
60+ HUBs
~2,000,000 Visitors Annually

hubzero

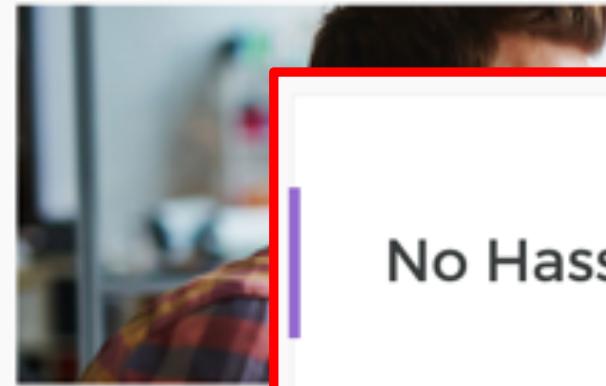


hubzero

- Disparate Sciences
 - Disparate Audiences
 - Disparate Schedules
 - Disparate Funding Sources
-
- Common Need for Framework
 - Common Desire to Share Competencies
 - Common Desire for Sustainable Partner

Operations

HUBzero Services



Bring your services
to the web

No Hassle Hub

- Cost effective
- Continuous upgrades
- Continuous security monitoring
- Continuous patching
- Dedicated community manager representative

[CHECK IT OUT](#)

No Hassle Hub

- Cost effective
- Continuous upgrades
- Continuous security monitoring
- Continuous patching
- Dedicated community manager representative

[CHECK IT OUT](#)

Open Source

- No cost
- Modify the code yourselves
- Host on your machine

[CHECK IT OUT](#)

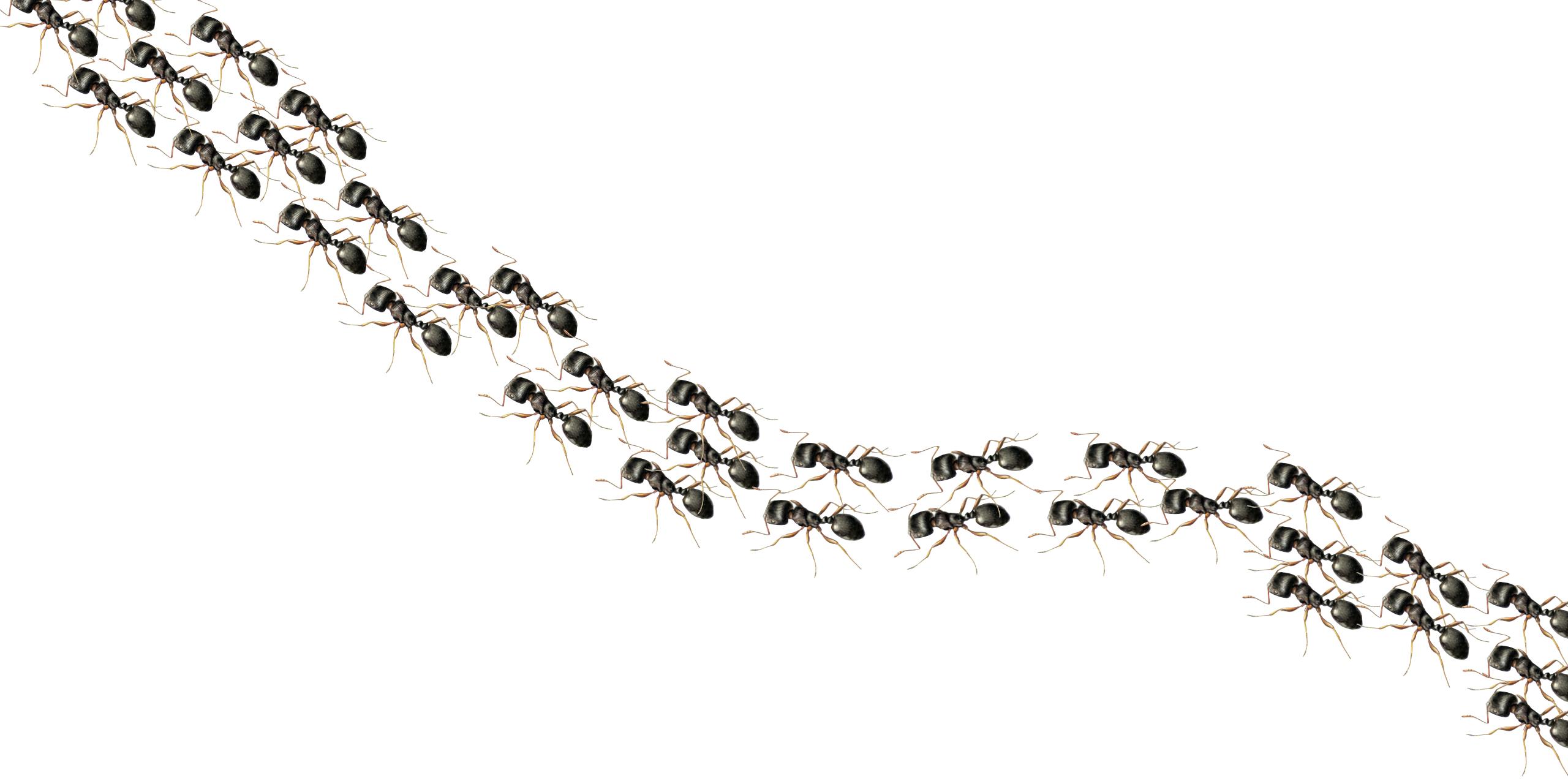
Less than 1/3 the cost of an FTE



Drawing by Wright



Drawing by Wright





<http://www.smartprint.net.au/2012/04/24/working-together-to-provide-managed-print-solutions/>

