

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



SGCI Incubator: Sustainability Support for Science Gateways

Sandra Gesing
sandra.gesing@nd.edu



Award Number
ACI-1547611

November 14, 2017
SC17, Denver, CO

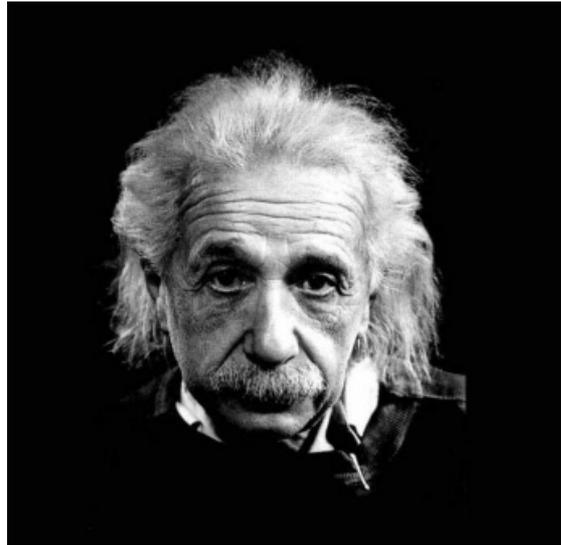


SC17
Denver, CO | hpc
connects.

State of the Art in Research

Increased complexity of

- research questions
- hardware
- software
- instruments
- data volume
- data formats

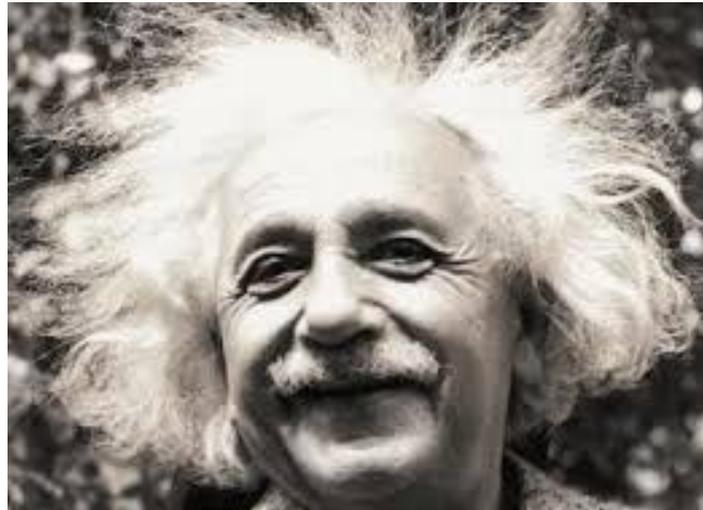


The need for **end-to-end solutions** for accessing **data, software, computing services, and equipment** specific to the needs of a science or engineering discipline

Science Gateways

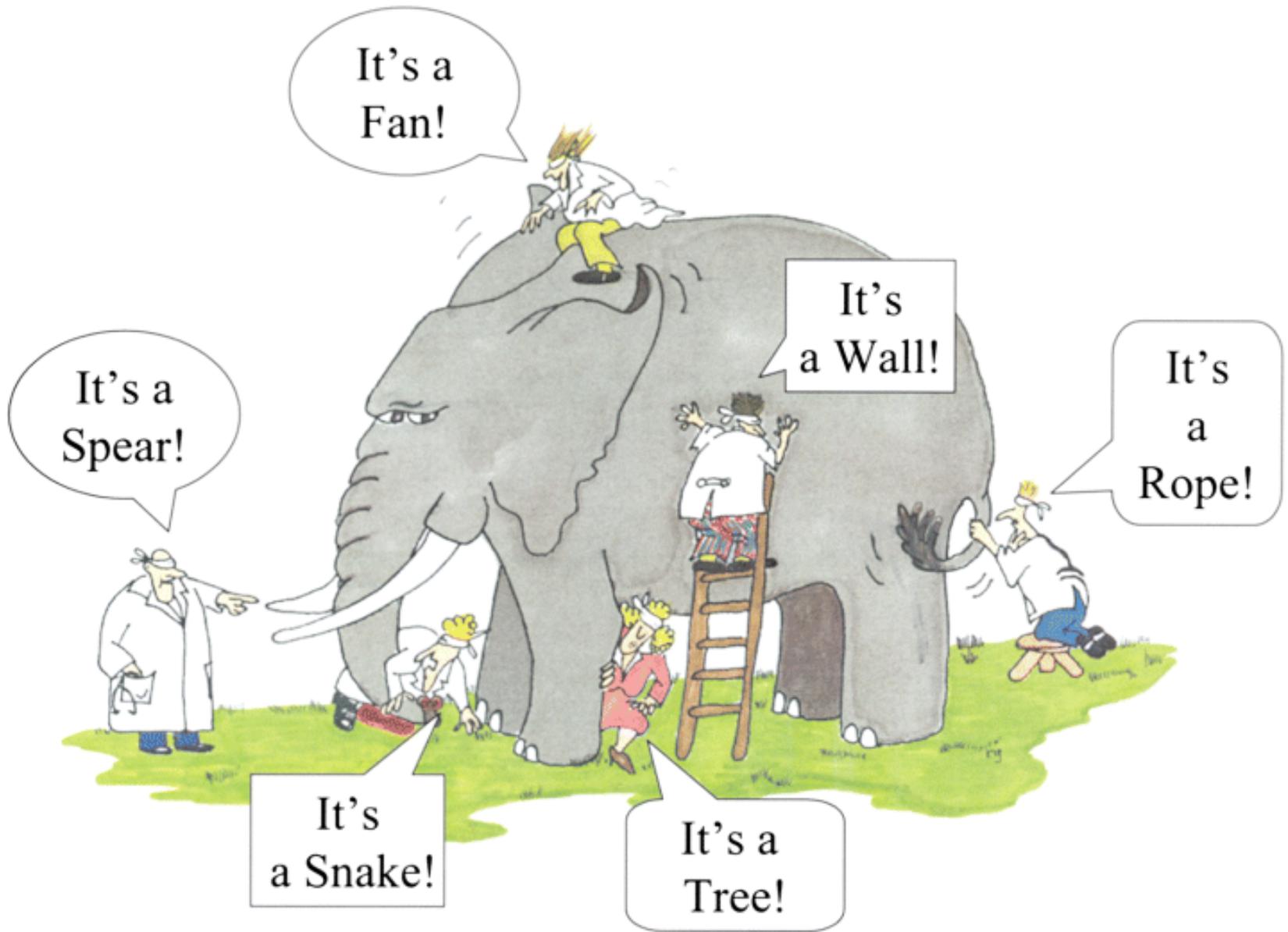
Increased complexity of

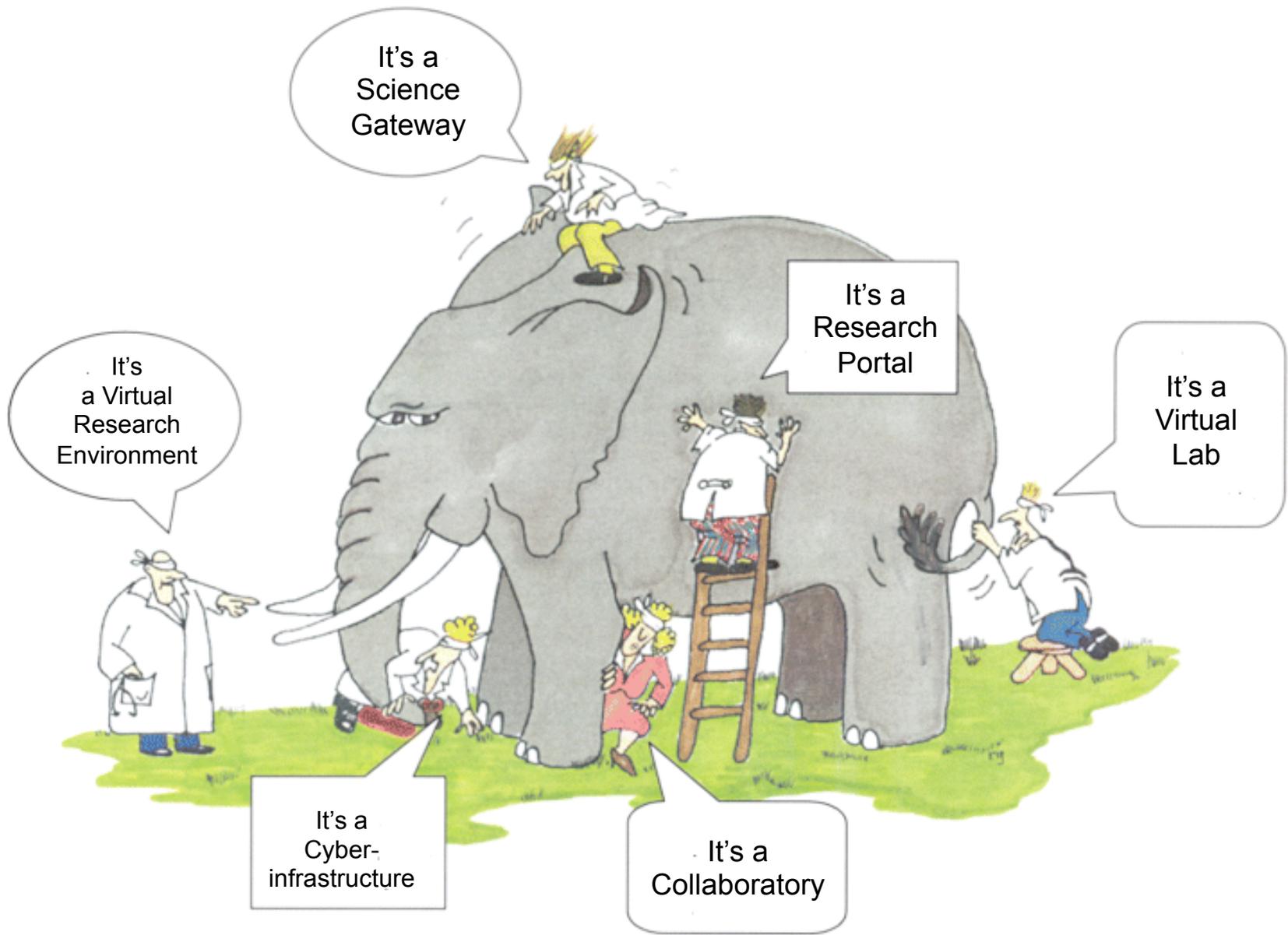
- research questions
- hardware
- software
- instruments
- data volume
- data formats



Science Gateways!

The need for **end-to-end solutions** for accessing **data, software, computing services, and equipment** specific to the needs of a science or engineering discipline





Science Gateways Survey 2014

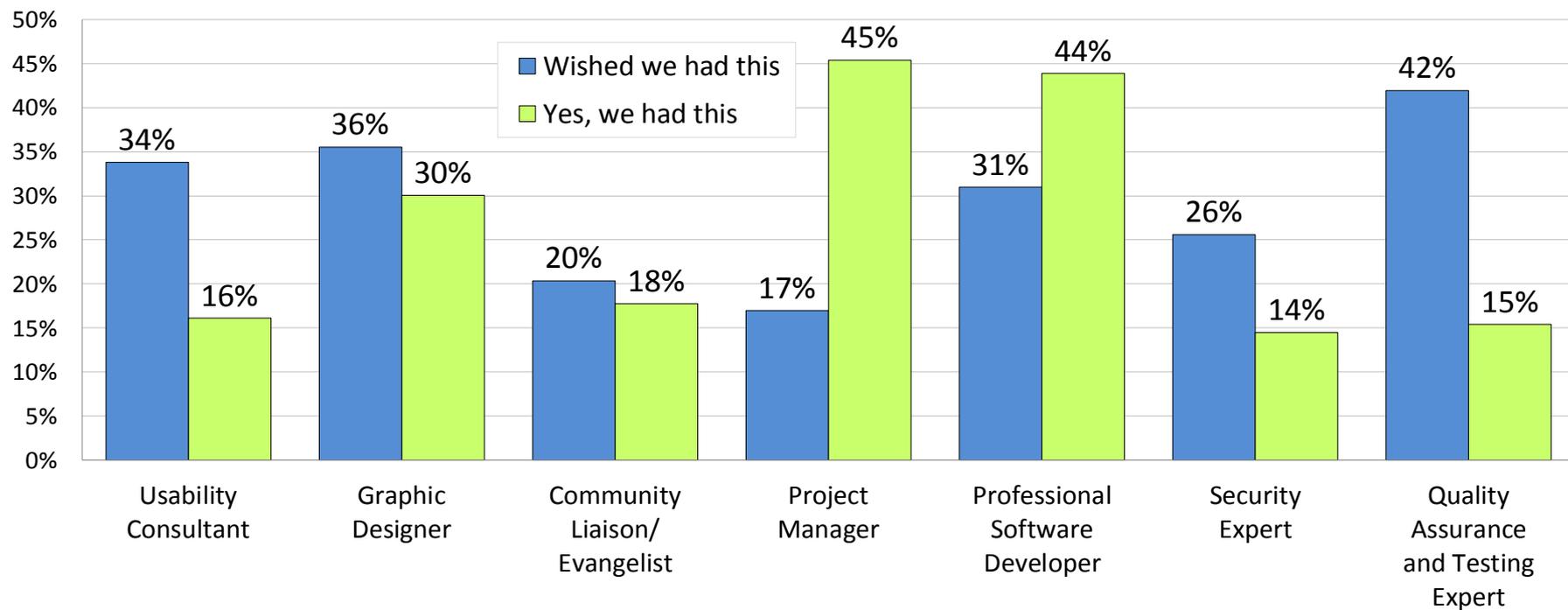
- sent out to 29,000 persons
- 4,957 responses from across domains
- 52% from life, physical or mathematical sciences
- 32% from computer and information sciences or engineering
- 45% develop data collections
- 44% develop data analysis tools

What services would be helpful?

Proposed Service	% Interest
Evaluation, impact analysis, website analytics	72%
Adapting technologies	67%
Web/visual/graphic design	67%
Choosing technologies	66%
Usability Services	66%
Visualization	65%
Developing open-source software	64%
Support for education	64%
Community engagement mechanisms	62%
Keeping your project running	62%
Legal perspectives	61%
Managing data	60%
Computational resources	59%
Mobile technology	59%
Database structure, optimization, and query expertise	59%
Data mining and analysis	58%
Cybersecurity consultation	57%
Website construction	57%
Software engineering process consultation	53%
Source code review and/or audit	51%
High-bandwidth networks	45%
Scientific instruments or data streams	44%
Management aspects of a project	38%

Science Gateways Survey 2014

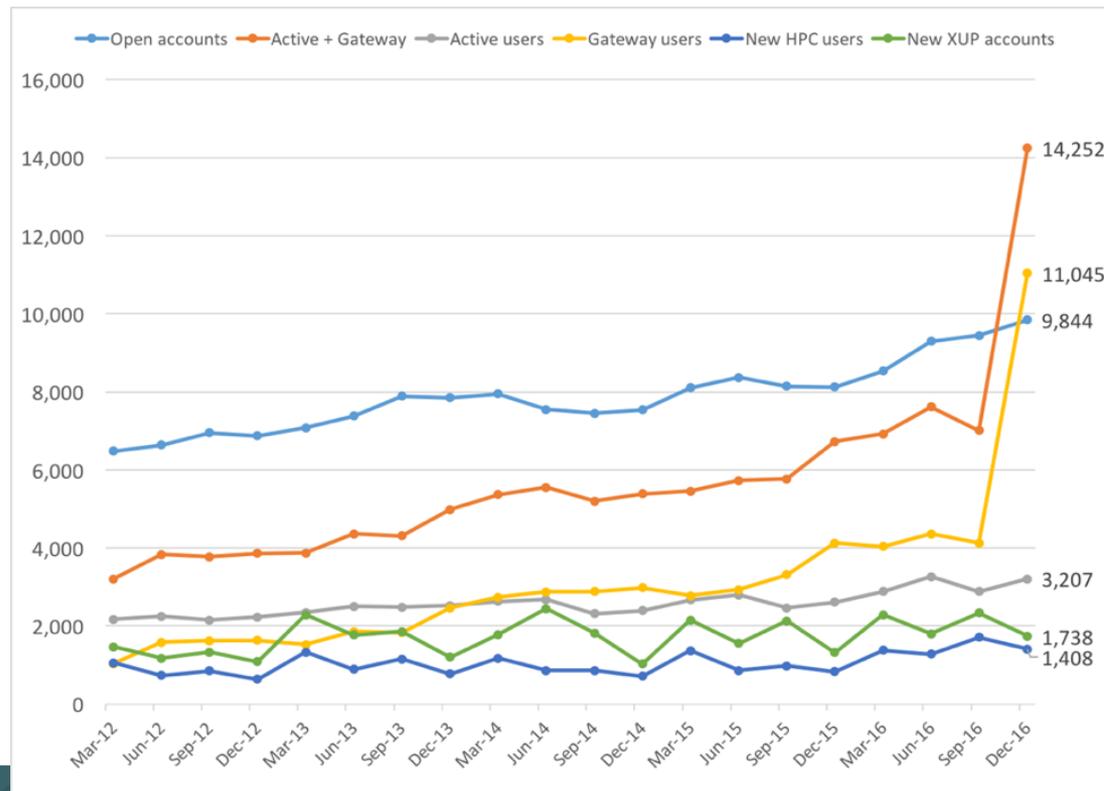
Well-designed gateways require a variety of expertise



Large Infrastructure Projects

- 2004 TeraGrid project director Rick Stevens recognized growth in scientific portal development and proposed the Science Gateway Program
- Followed up by XSEDE's Science Gateway Program

Gateway users are 77% of active XSEDE users in Q4 2016

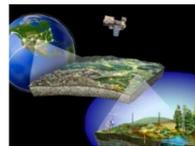


Funding Bodies

Funding bodies such as NSF and NIH mention science gateways direct in solicitations and roadmaps!

OAC supports Research Cyberinfrastructure to uniquely enable collaboration and discovery frontiers at all scales

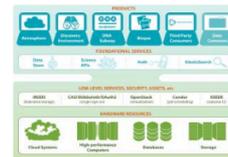
Shared resources, capabilities & services across the scientific workflow



CI-Enabled Instrumentation



Computing Resources



Gateways, Hubs, and Services



Cloud Resources & Services



Data Networks, Cybersecurity



Coordination & User support



Software, Applications, Workflow Systems

<https://dibbs17.org/report/Presentations/KeynoteQualters.pdf>

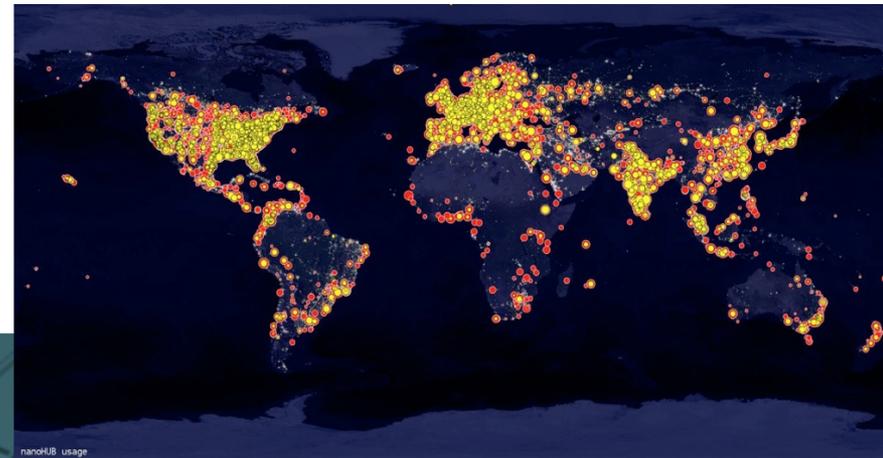
Technologies

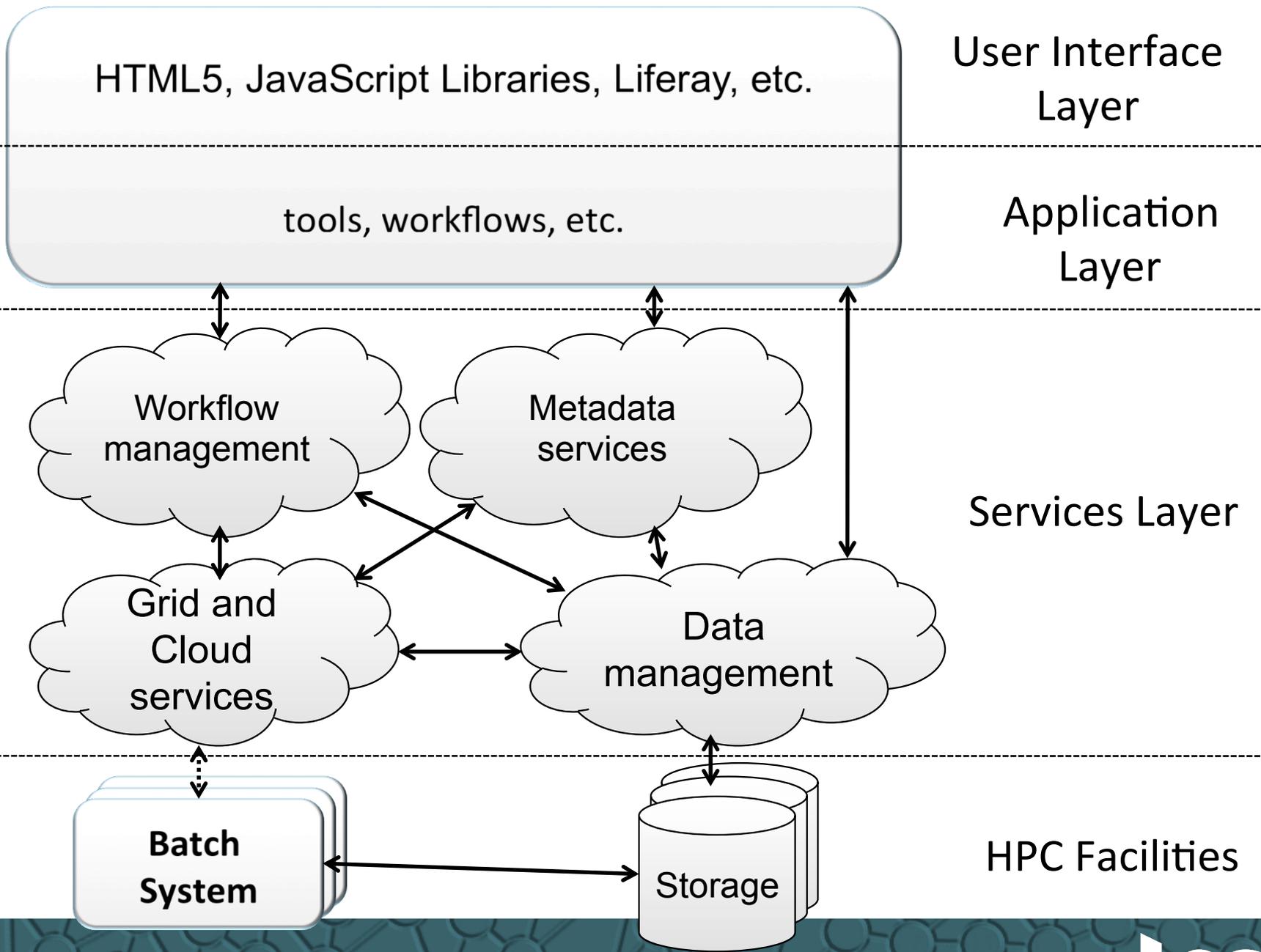
- Widely used complete frameworks (Galaxy, HUBzero, Open Science Framework, Globus etc.)
- RESTful APIs and support of multiple programming languages in widely used frameworks (Apache Airavata, the Agave platform, etc.)
- Reused interface implementations such as the one of CIPRES with its RESTful API (CIPRES has served more than 20,000 users to date)
- Science gateways as a service with provision of hardware in the background such as SciGap (Science Gateway Platform as a Service)

Lessons learned: approaches should be technology agnostic, using APIs and standard web technologies OR deliver a complete solution

Community Engagement is key

HUBzero users world wide





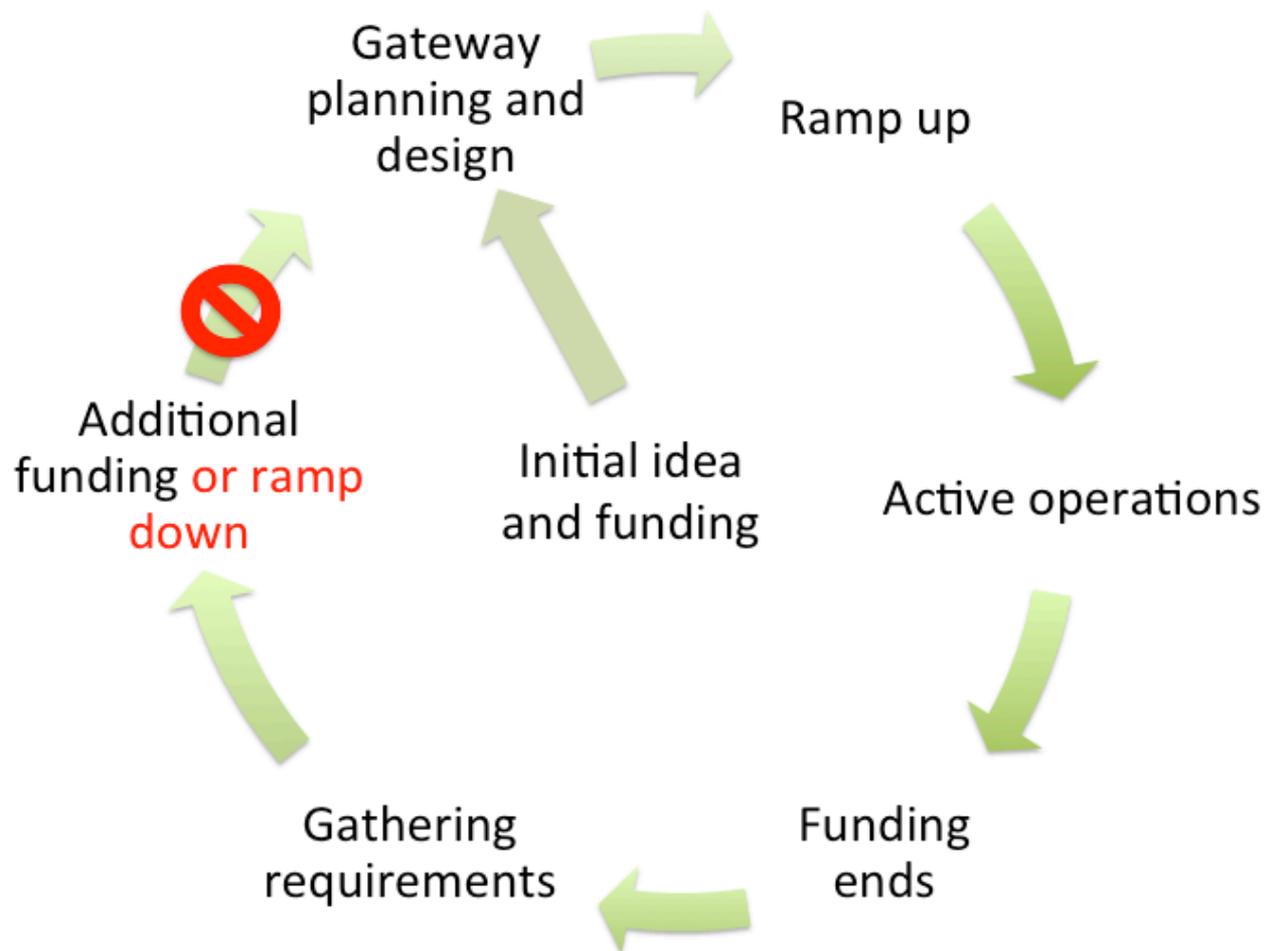
User Interface Layer

Application Layer

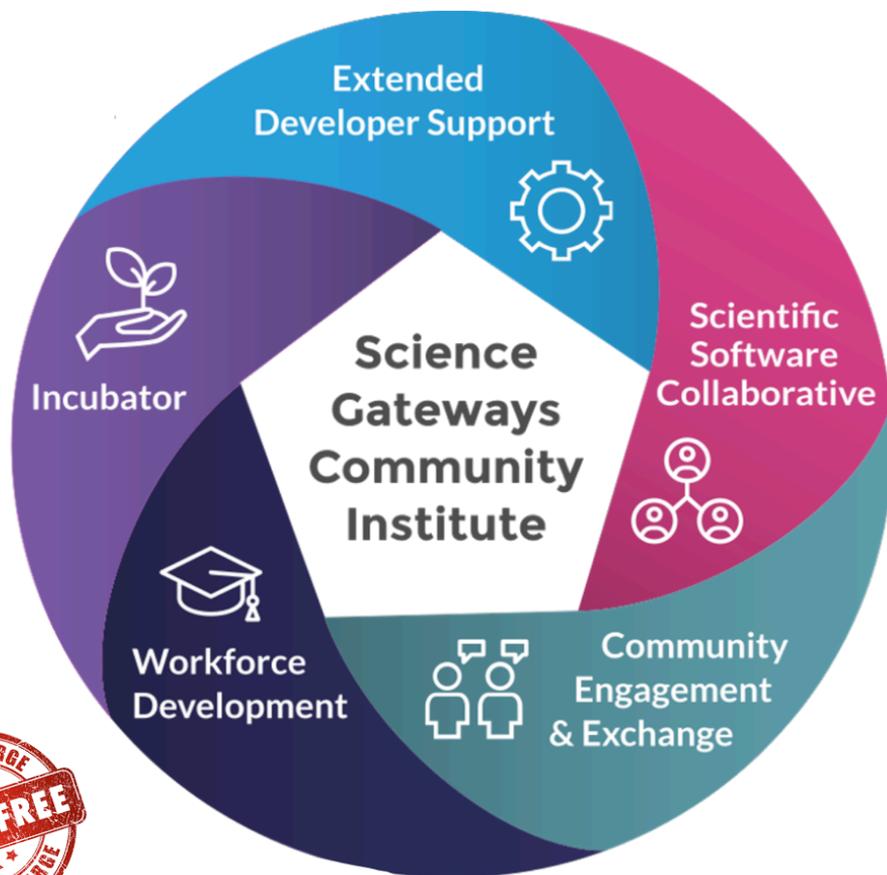
Services Layer

HPC Facilities

Lifecycle of a Science Gateway



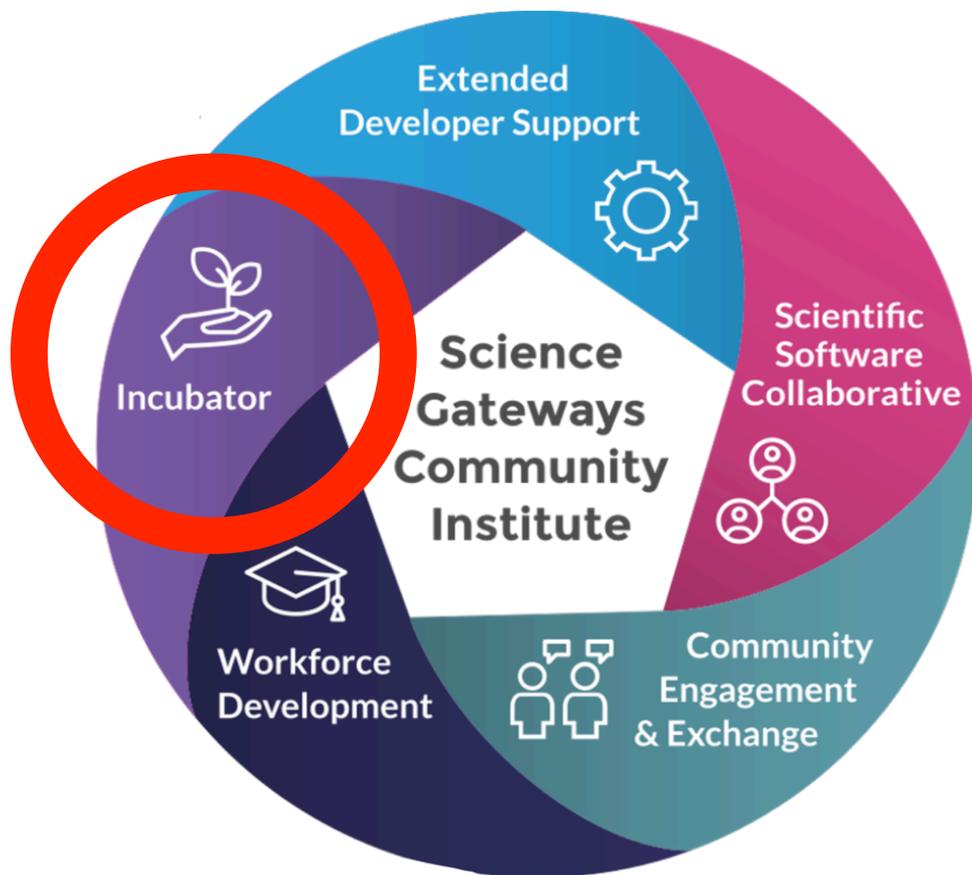
Science Gateways Community Institute



help@sciencegateways.org
<http://sciencegateways.org/>

- Diverse expertise on demand
- Longer term support engagements
- Software and visibility for gateways
- Information exchange in a community environment
- Student opportunities and more stable career paths

Science Gateways Community Institute



- Diverse expertise on demand
- Longer term support engagements
- Software and visibility for gateways
- Information exchange in a community environment
- Student opportunities and more stable career paths

help@sciencegateways.org
<http://sciencegateways.org/>

Incubator Service

A Framework for Decision Making

Technology Planning

- Choosing technologies
- Cybersecurity
- Software engineering
- Interfaces to compute and data

Business Planning

- Business model development
- Financial planning
- Project management
- Software licensing
- Staff and sustainability planning

Client Interaction Planning

- Usability studies
- Web/visual/graphic design
- Impact measurement
- Community engagement
- Support for education

Specialized Expertise

Security

- Center for Trustworthy Scientific Cyberinfrastructure

Sustainability

- Nancy Maron, creator of the ITHAKA S+R course on Sustaining Digital Resources

Evaluation & Impact Measurement

- Ann Zimmerman Consulting

Campus Resource Development

Network / Cohort Formation

Common Experiences

- Training sessions
- Group interactions

Continuing Engagement

- Customized structure, content, goals
- Mentoring
- Pay It Forward

An Ongoing
Dispassionate Ear

I have an idea!



Articulate the value of your gateway and how it's distinctively different from what already exists.

Who benefits?



Identify audience and stakeholder groups and consider how they impact your success.

Where does it fit in?



Establish where your gateway solution fits within the existing market landscape of partners and competitors.

How do I make it happen?



Define measurable goals for success and sustainability. Consider multiple needs such as technology, security, project management, usability, and funding.

How do I sell it?



Spread the word! Plan how to tell the unique story of your gateway.

Bootcamp at a Glance

- 5 full days
- Teams on science gateways
- Knowledge dissemination
- Interactivity
- Community formation
- Putting away the normal daily routine
- Homework

COURSE PLAN

MONDAY

TUESDAY

WEDNESDAY

THURSDAY

FRIDAY

Your Audience(s)
and Stakeholders

Technology,
Open Source

Cybersecurity

Outreach

Mapping the
landscape

User Centered
Design

On-Campus
Support

Your
Sustainability
Model – Making
Your Case

Introductions

Course Outline
and Goals

The “Napkin”
Drawing

Defining Your
Value
Proposition

Goal Setting

Funding
Models

Impact
Measurement
Case Study

Going beyond
your initial
market

Budgeting

Marketing &
Outreach

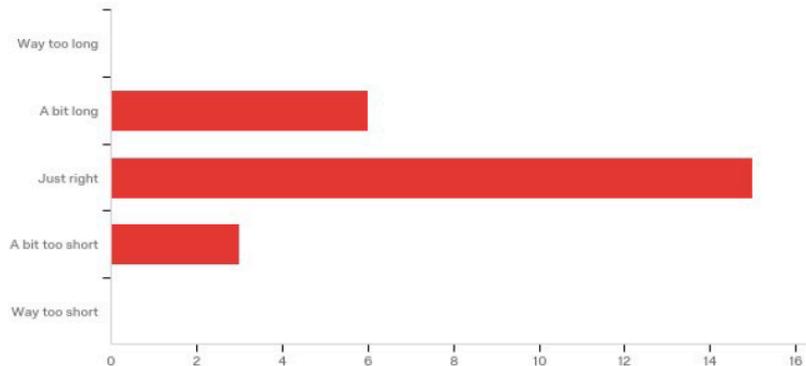
Next Steps

Closing

Bootcamps

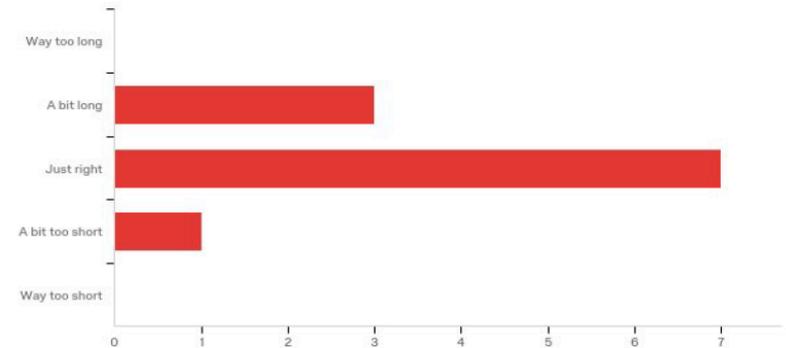
- twice per year
- additional ones can be booked (travel expenses for presenters)
- adapted to feedback

Q6 - How would you rate the length of the Bootcamp?



Bootcamp 1

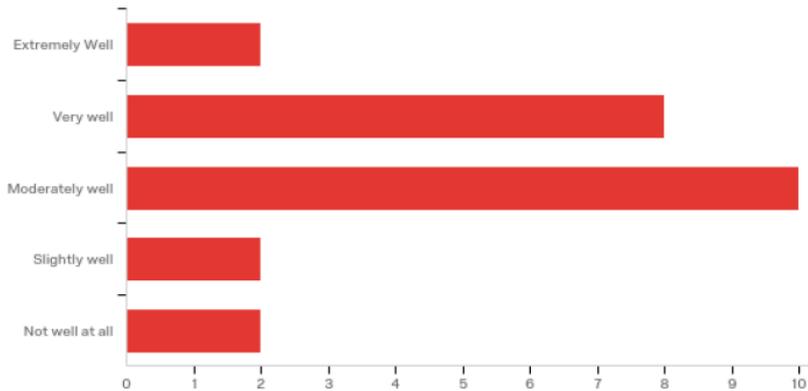
Q6 - How would you rate the length of the Bootcamp?



Bootcamp 2

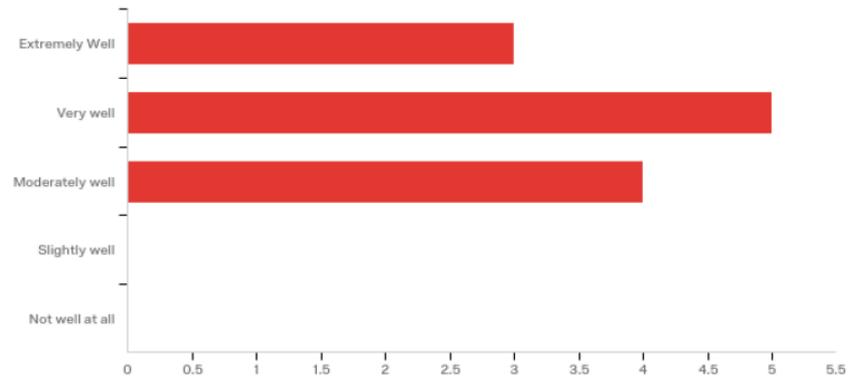
Feedback to the Bootcamps

Q4 - How well did the information you received prior to the Bootcamp prepare you to participate?



Bootcamp 1

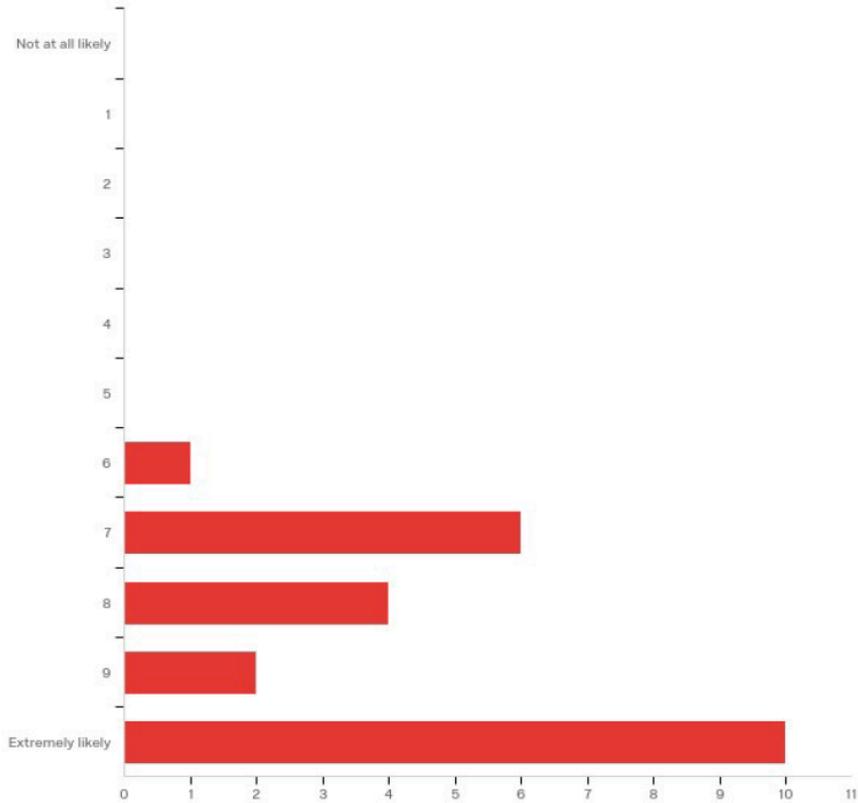
Q4 - How well did the information you received prior to the Bootcamp prepare you to participate?



Bootcamp 2

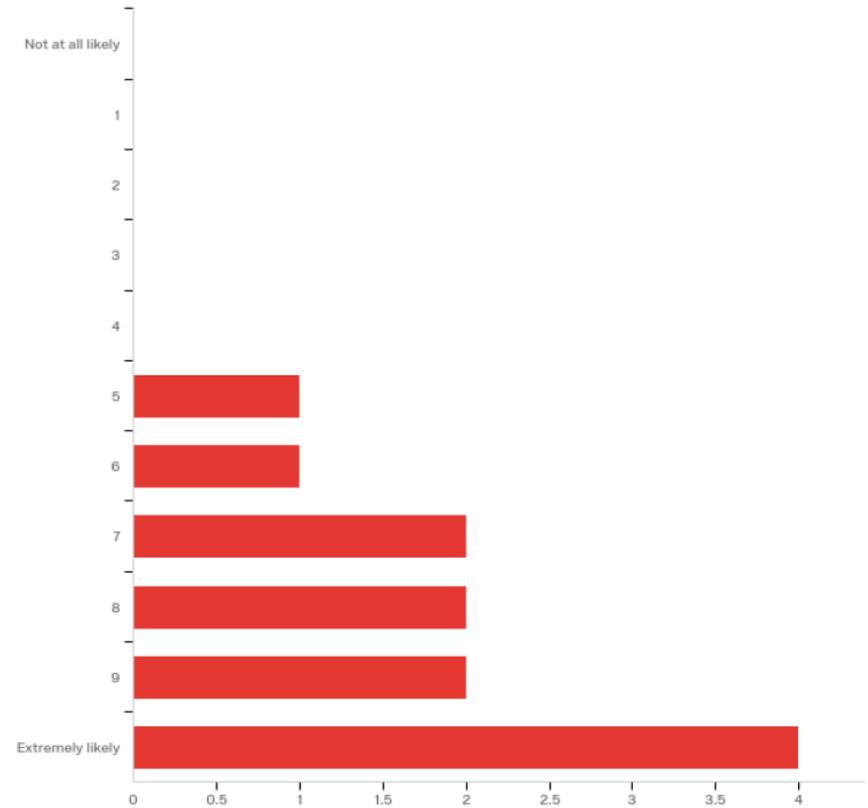
Feedback to the Bootcamps

Q8 - How likely is it that you would recommend this workshop to a colleague?



Bootcamp 1

- How likely is it that you would recommend this workshop to a colleague?



Bootcamp 2

Feedback to the Bootcamps

"The Science Gateways Bootcamp is useful to people involved at any stage of their science gateway project, and in any role. Their experts have you consider all aspects of sustainability, from understanding your unique value proposition and audience, through to best practices and standards in software development. They cultivate a lively, active, hands-on environment. A great experience overall."

Steve Androulakis
Manager — Community Platforms,
NeCTAR.org.au, ANDS.org.au, RDS.org.au
(now at Amazon)

"The SGCI Bootcamp has fundamentally altered, in a constructive way, my view toward developing and implementing the projects on which I work. The work during the week helped me to generate a new approach to the gateway on which I'm working that has much greater potential to be successful, and I intend to apply the ideas and concepts from the Bootcamp to all project phases, from idea to implementation."

Christopher Lenhardt
Domain Scientist for Environmental Data Science &
Systems, RENCI Data Observatory
Risk Analytics Discovery Environment (RADE) gateway

<https://sciencegateways.org/-/reflections-from-the-inaugural-science-gateways-bootcamp-in-april-2017>

Sustainability via On-Campus Teams

- Great visibility for the institution's research activities
- Synergy effects between projects
- Shared resources, costs and expertise across departments
- Lower learning curves
- Expertise that is otherwise difficult for individual projects to obtain
- Ability to retain top-quality research computing support by providing interesting projects

Success Stories

Many universities have successful centers/groups with centralized services for science gateways

- Center for Research Computing at University of Notre Dame
- HUBzero[®] Team at Purdue University
- Science Gateways Research Center at Indiana University
- Science Gateway Group at TACC at the University of Texas, Austin
- ...

ND CRC in 2006-2008

- Effort with 7 FTEs centrally funded
- HTC and HPC Computing and basic user support
 - One centrally funded cluster plus multiple faculty funded clusters in various cabinets on campus
 - Around 300 active users
 - 80% centrally funded hardware
- No other kinds of research computing services
- Underserved social sciences and humanities for their need on science gateways
- ~1000 faculty, ~12000 students

ND CRC Vision

To become an internationally recognized multidisciplinary research computing center based upon our reputation for facilitating and accelerating discovery through effective and novel applications of cyberinfrastructure.

ND CRC Director's thoughts on Vision

- “Nice vision, but how we get there?”
- “What should we do first?”
- “Users should tell us what they need...”
- “HPC works fine, so let's focus on CI Development...”
- “We need portals and other collaborative environments”
- “We need good user support, and good, straight resource usage policies”
- “We need CI and HPC teams working together”
- “How do we fund all these with very limited resources?”
- “What is available out there?”
- And so on...

Implementation by Jarek Nabrzyski

Take the risk and hire people first

→ train people

→ generate / bring projects

→ assign people to projects and focus on getting more projects and more people -> etc...

→ if not successful then exit, i.e. quit 😊

Fortunately, we had many projects, mostly science gateway projects of various kinds and difficulty.

ND CRC in 2017

- 45 FTEs with 70% of the staff salaries supported through grants and services
- HTC and HPC Computing and user support
 - 26,000 cores
 - 1,800 active users
 - 10 HPC engineers and user support staff
 - 30% of compute nodes are centrally funded
- Cyberinfrastructure development
 - ~15-20 CI projects each year with ~35 faculty from various departments including social sciences and humanities
 - supported by ~15 research programmers, ~8 computational scientists, some FTE fractions of HPC engineers, and a few (7) grad students and undergraduate interns (4-6)
- ~1000 faculty, ~12000 students

HUBzero History: Operating Model Path

- 1996: PUNCH (precursor to nanoHUB)
- 2002: NSF funding for nanoHUB
- 2007: Spin-out of HUBzero from nanoHUB
- 2015: Diversification, self-funded sustainability
- Composition:
 - Hubs hosted by Purdue (for a variety of institutions)
 - Foundation members running their own hubs
 - Open source hubs

HUBzero History: Key Team Enablers

- 1996: Vision (Mark Lundstrom)
- 2002: Vertical Expansion (Gerhard Klimeck)
- 2007: Horizontal Expansion (Michael McLennan)
- 2015: Business Transition (Michael Zentner)
- Composition:
 - Hubs hosted by Purdue (for a variety of institutions)
 - Foundation members running their own hubs
 - Open source hubs

HUBzero in 2017

- A foundation, a group at Purdue, and a software framework
- 25 full time software professionals (+5 vacancies) specializing in:
 - Cybersecurity
 - Web programming
 - User experience design
 - Scientific application development
 - Analytics
 - Middleware
 - High performance computing
 - System administration
 - Customer service
- Servicing over 2 million total visitors annually
- Entirely self funded

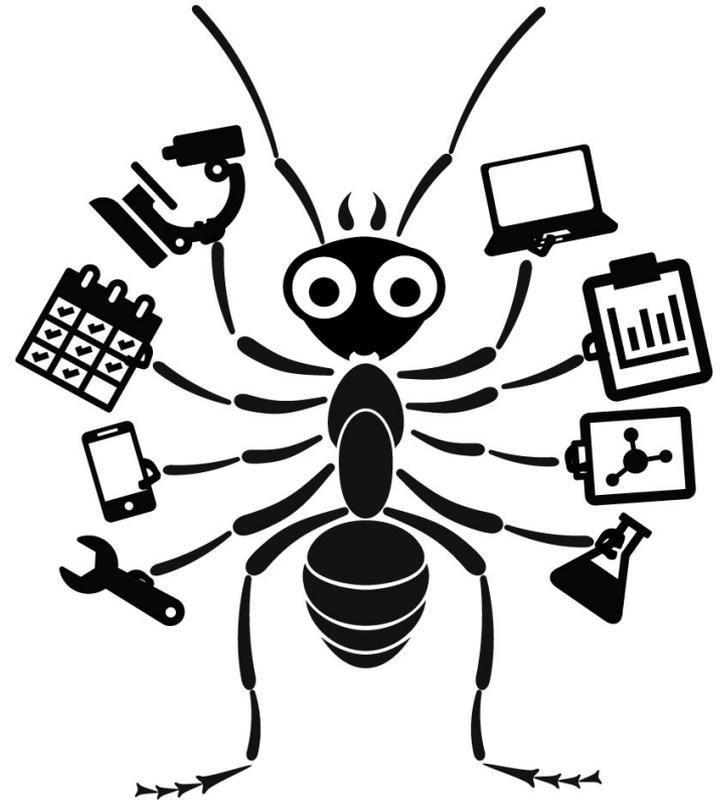
Even with success we must continue to innovate new business models, partnerships, delivery capabilities, and service architectures.

On-Campus Teams - Opportunities

- A breadth of interesting topics in the science gateway creation process
 - Novel frameworks and web applications
 - Inter- and multidisciplinary work
 - Contributing to grand challenges, e.g., Malaria eradication
 - ...
- A breadth of interesting roles
 - Designers, statisticians, librarians, machine learning experts, usability experts,...

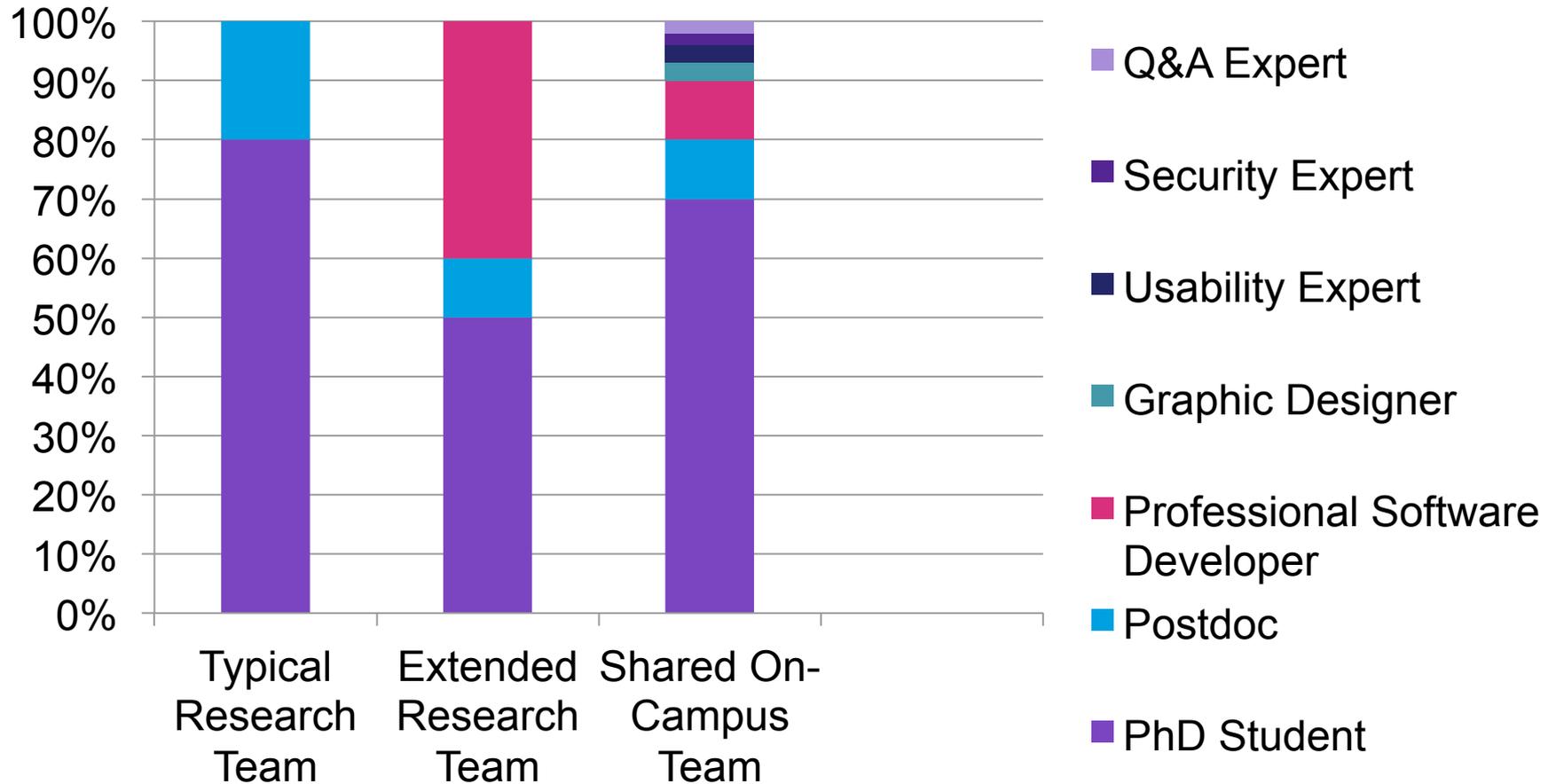
On-Campus Teams - Challenges

- Some topics and roles are only needed for part of the project
 - ⇒ Not fully funded positions via one project
 - ⇒ For diverse expertise, several different people are needed



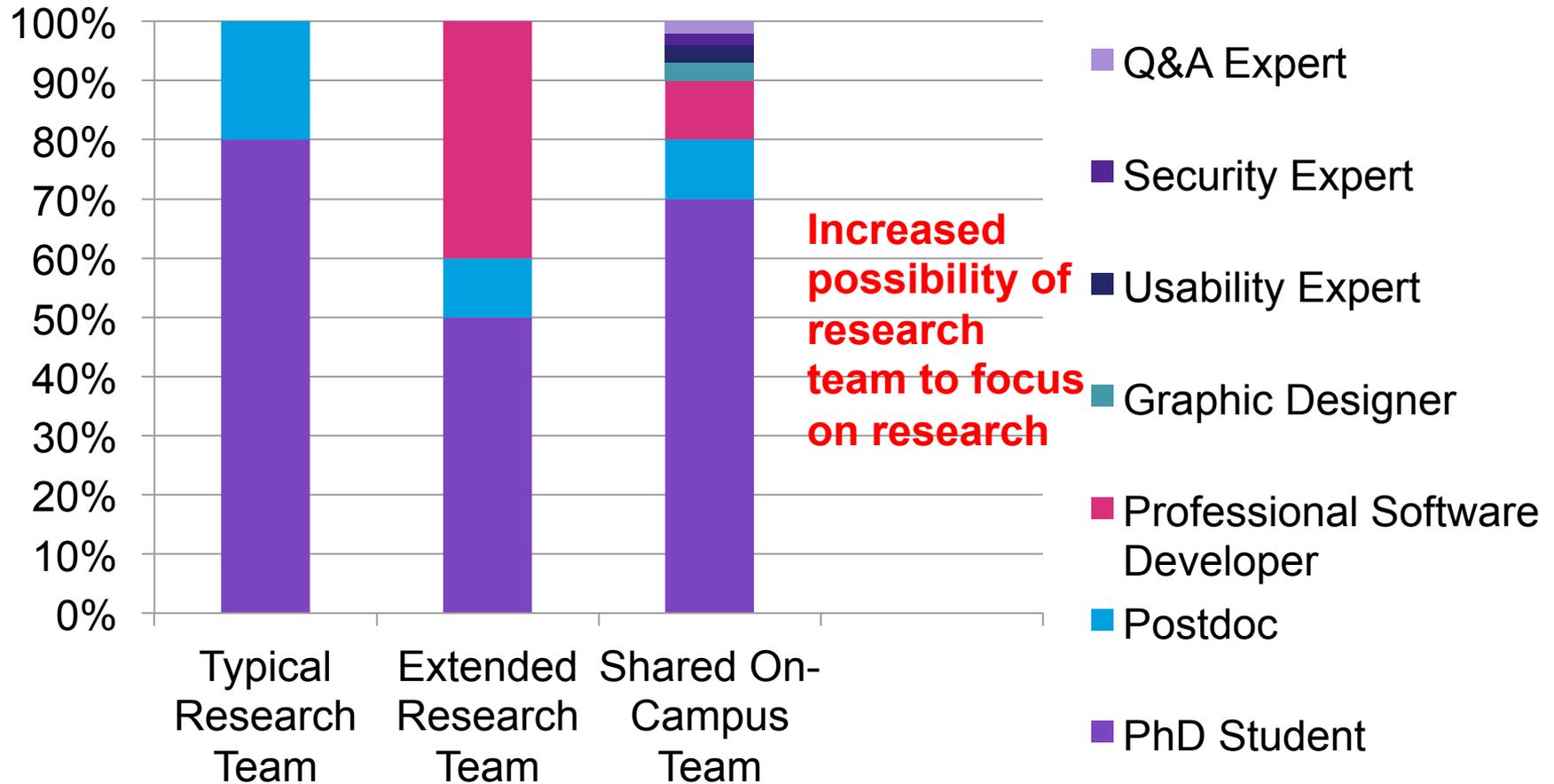
Sustainability via On-Campus Teams

Potential salary cost distribution



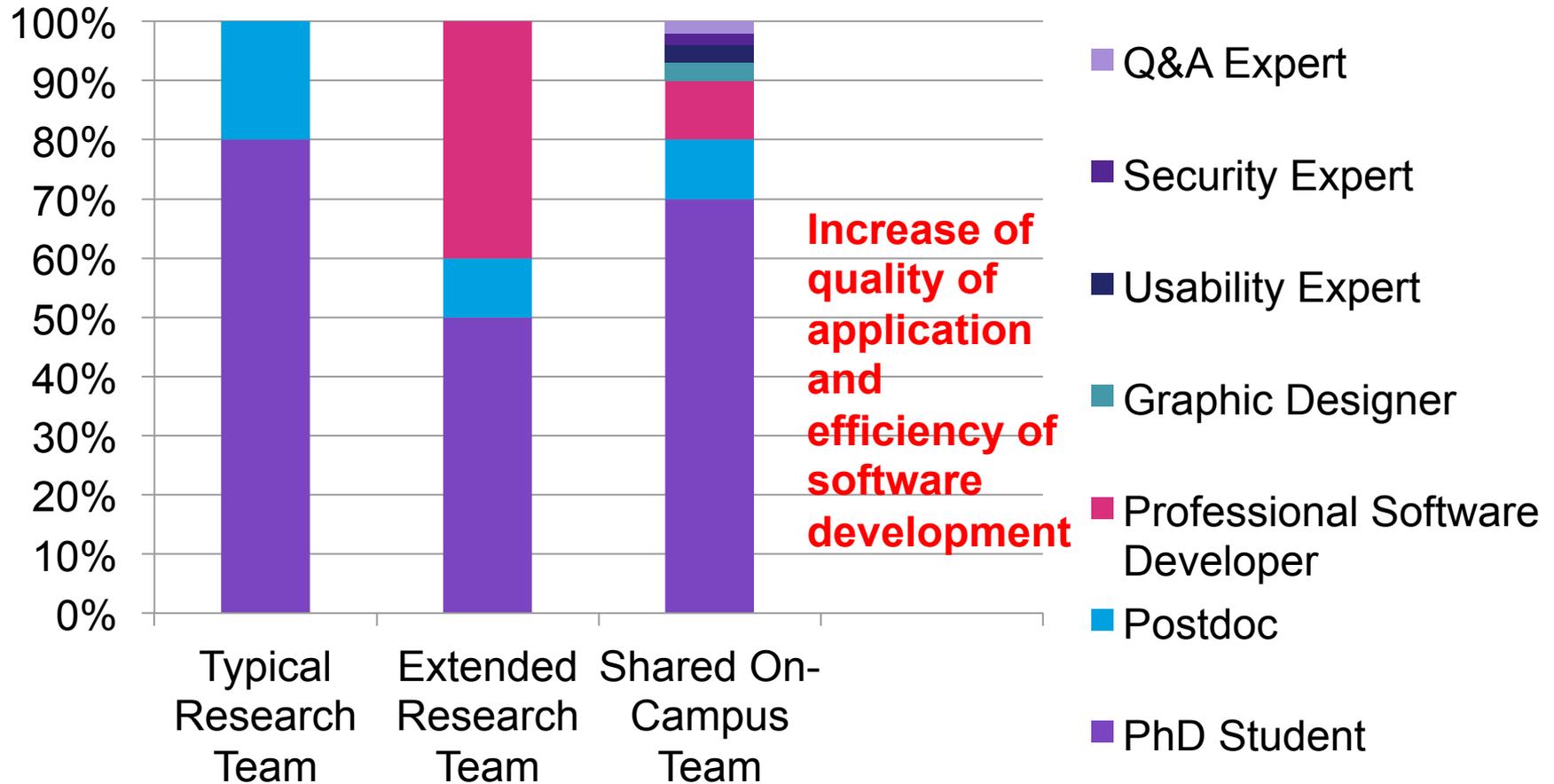
Sustainability via On-Campus Teams

Potential salary cost distribution



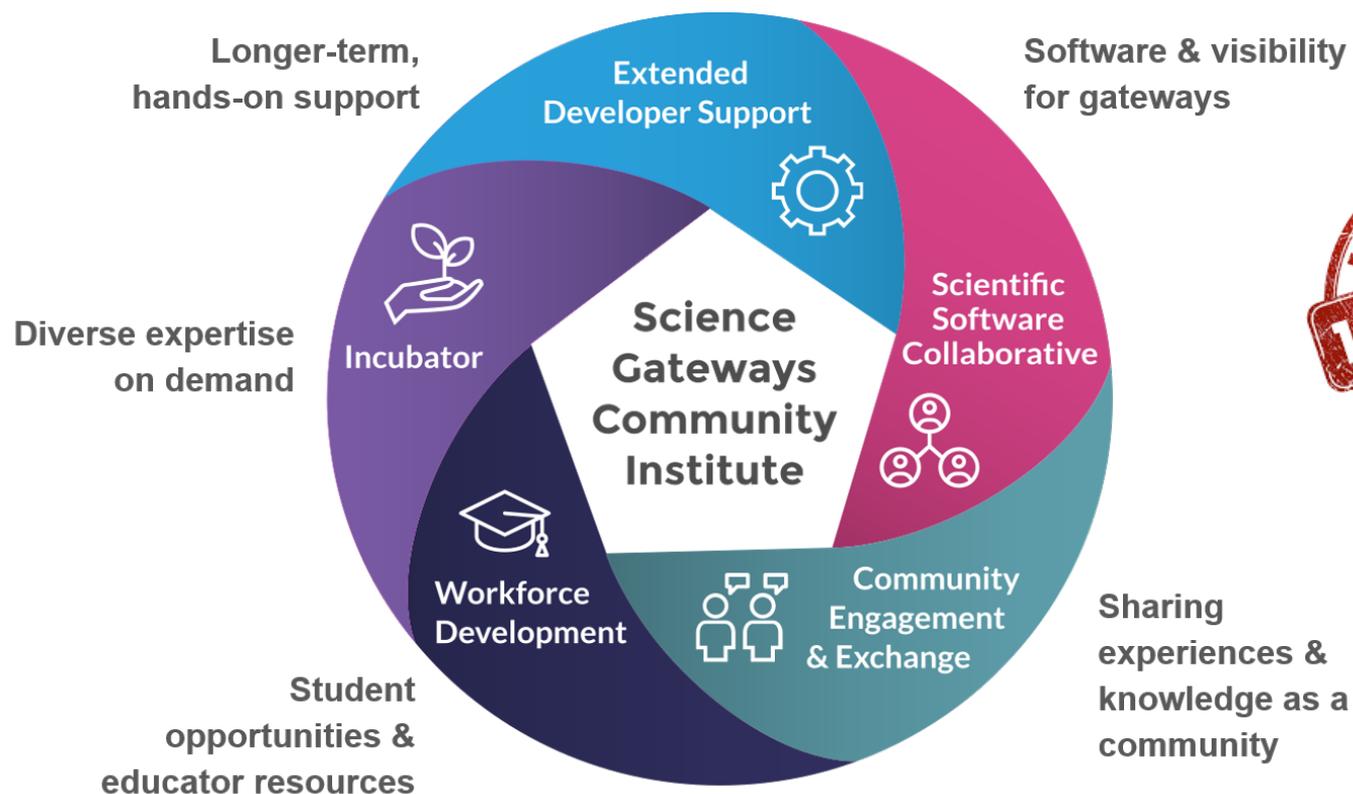
Sustainability via On-Campus Teams

Potential salary cost distribution



Funding Mechanisms – External Resources

Now is the right time – Get support from SGCI via diverse services!



Funding Mechanisms – External Resources

- Get support from SGCI via diverse services
- Contractors from mature science gateway frameworks

⇒ Means to create and/or maintain science gateway(s) successfully while working on building up internal resources

Funding Mechanisms – Internal Resources

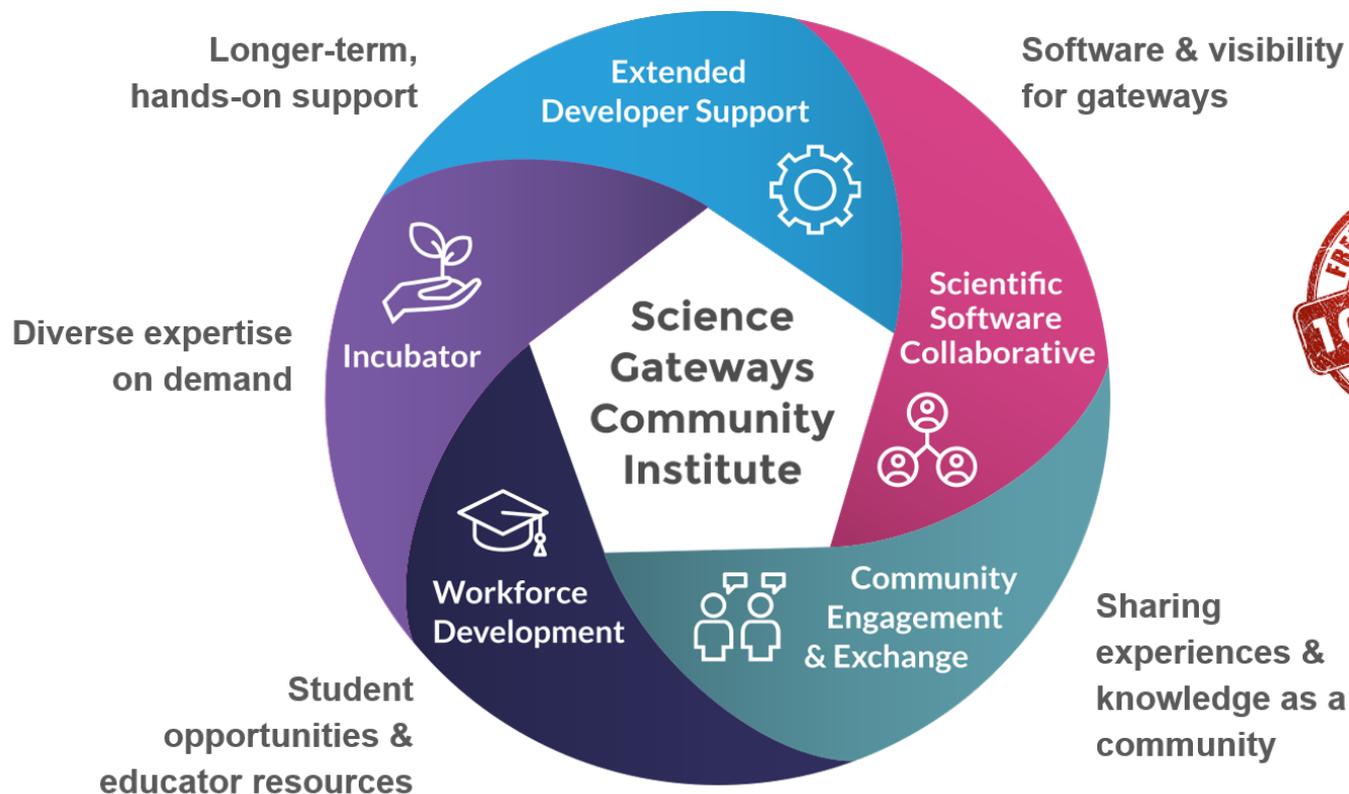
- Funding via involvement of some person months in grants
- Funding on some hard money from universities
- Funding via re-charge
- Funding via NSF for building cyberpractitioner career-paths (under development - models like full funding the first year and increasing responsibility on the side of universities)

Using Existing Free On-Campus Resources



- Do you have people such as “digital librarians”?
They are generally not only serving humanities and have great knowledge about data preservation, data lifecycle, programming skills, ..
- Do you have data scientists?
They probably know about machine learning, meta-data, ontologies, statistics ...
- Do you have an HPC center?
They know how to access HPC resources, VMs, containerization, distributed data management ...

Thanks for your attention!



help@sciencegateways.org
<http://sciencegateways.org/>