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

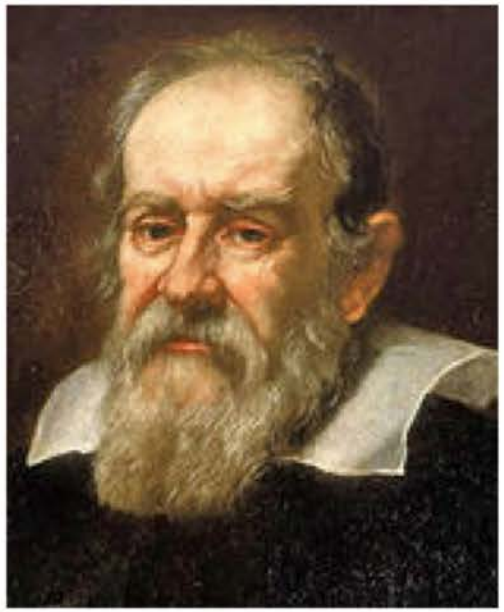
SGCI

Science Gateways Community Institute

Nancy Wilkins-Diehr
wilkinsn@sdsc.edu

Award Number
ACI-1547611

Science Gateways
First conceived in the 1600s!



"In the future, there will be opened a gateway and a road to a large and excellent science into which minds more piercing than mine shall penetrate to recesses still deeper."

– Galileo Galilei

Today's modern implementation

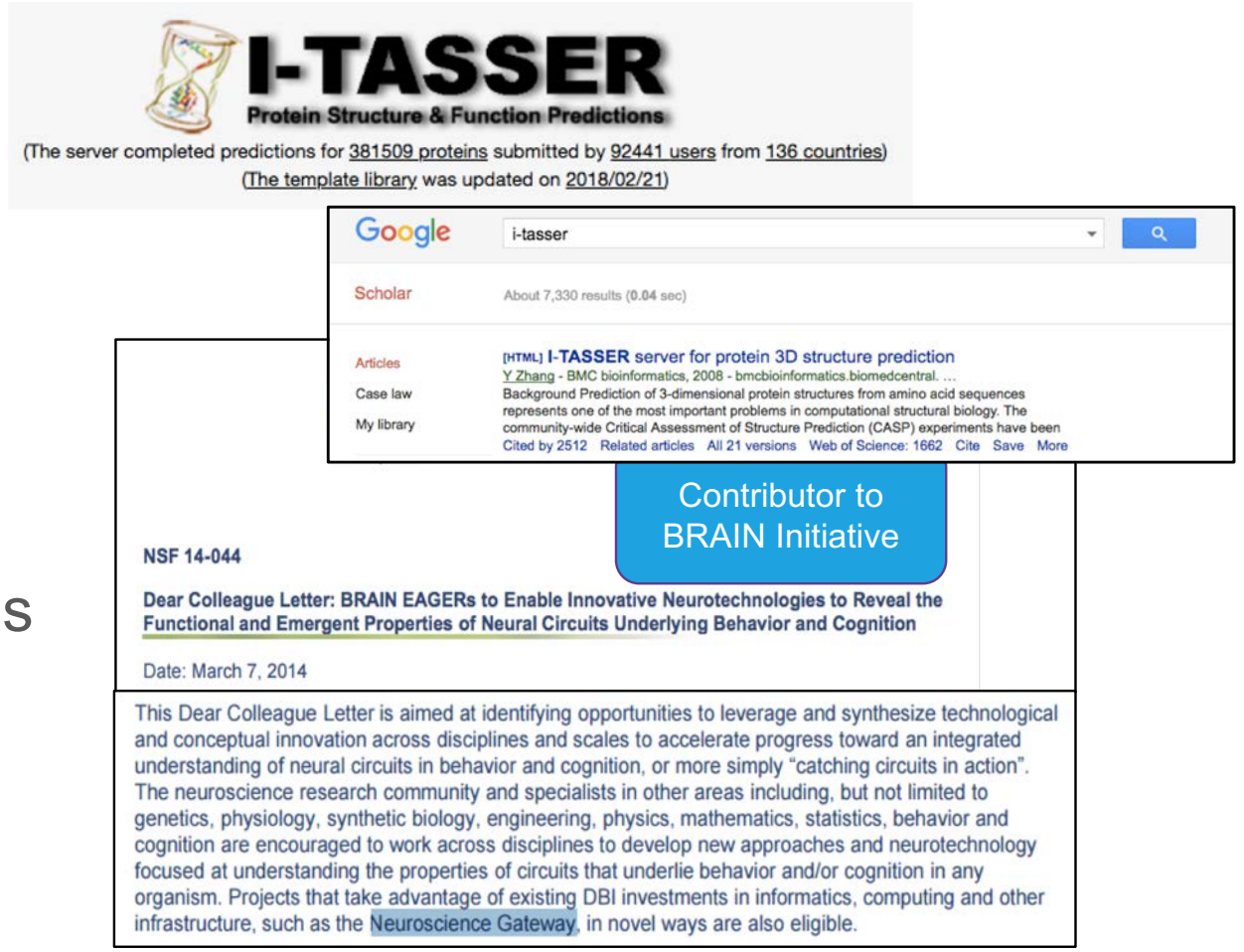
science gateway /sī′ əns gāt′ wā′/ *n.*

1. an online community space for science and engineering research and education.
2. a Web-based resource for accessing data, software, computing services, and equipment specific to the needs of a science or engineering discipline.

Gateways, research portals, virtual research environments, research platforms, virtual labs,…

Examples of how gateways are changing the conduct of science

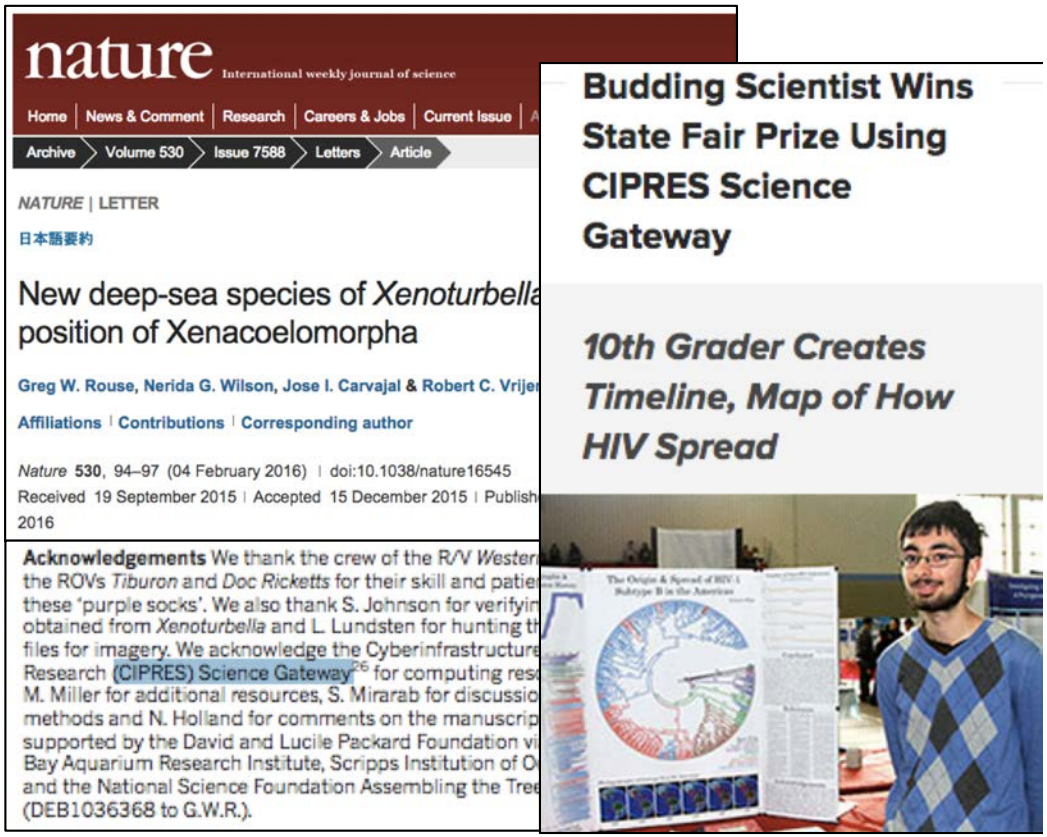
- Used by thousands of researchers, scholars and students every day
  - In all fields
- Citations are rising rapidly
- Named in solicitations as a launch pad to accelerate research



CIPRES science gateway

Premier researchers and high school students using the same tools

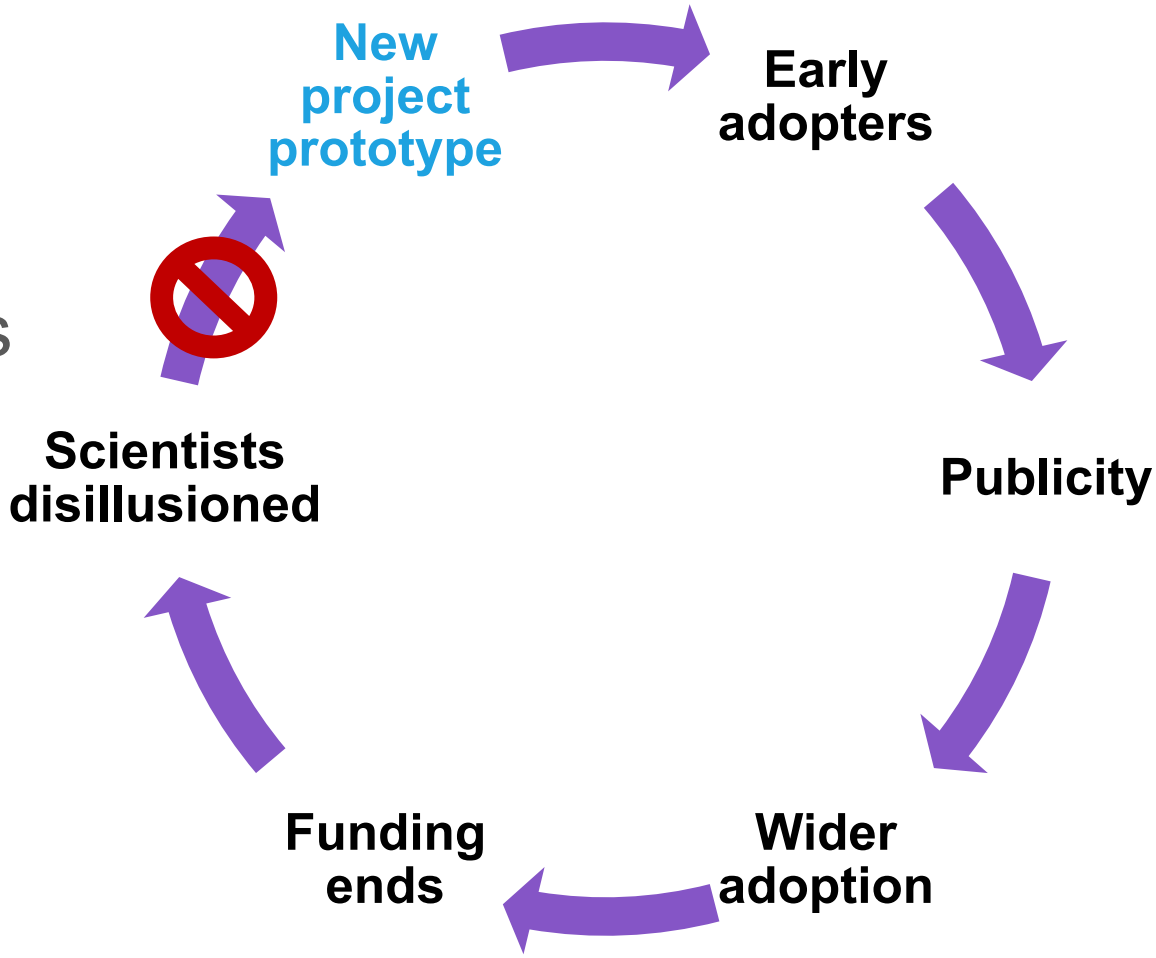
- 298 US research universities
  - From top-tier research institutions to non-PhD granting colleges, all EPSCOR states
- 6 K-12 school systems
- 43 non-governmental organizations
  - Museums including the Smithsonian Institution, the American Museum of Natural History, and the Field Museum
  - Botanical gardens, (e.g. Chicago, Rancho Santa Ana, and New York)
  - Institutes (e.g. JCVI and Broad)
- 10 US governmental agencies
  - Including NIH, USDA, NOAA, US Forest Service
- Curriculum delivery (84)
- 4000+ publications since 2010



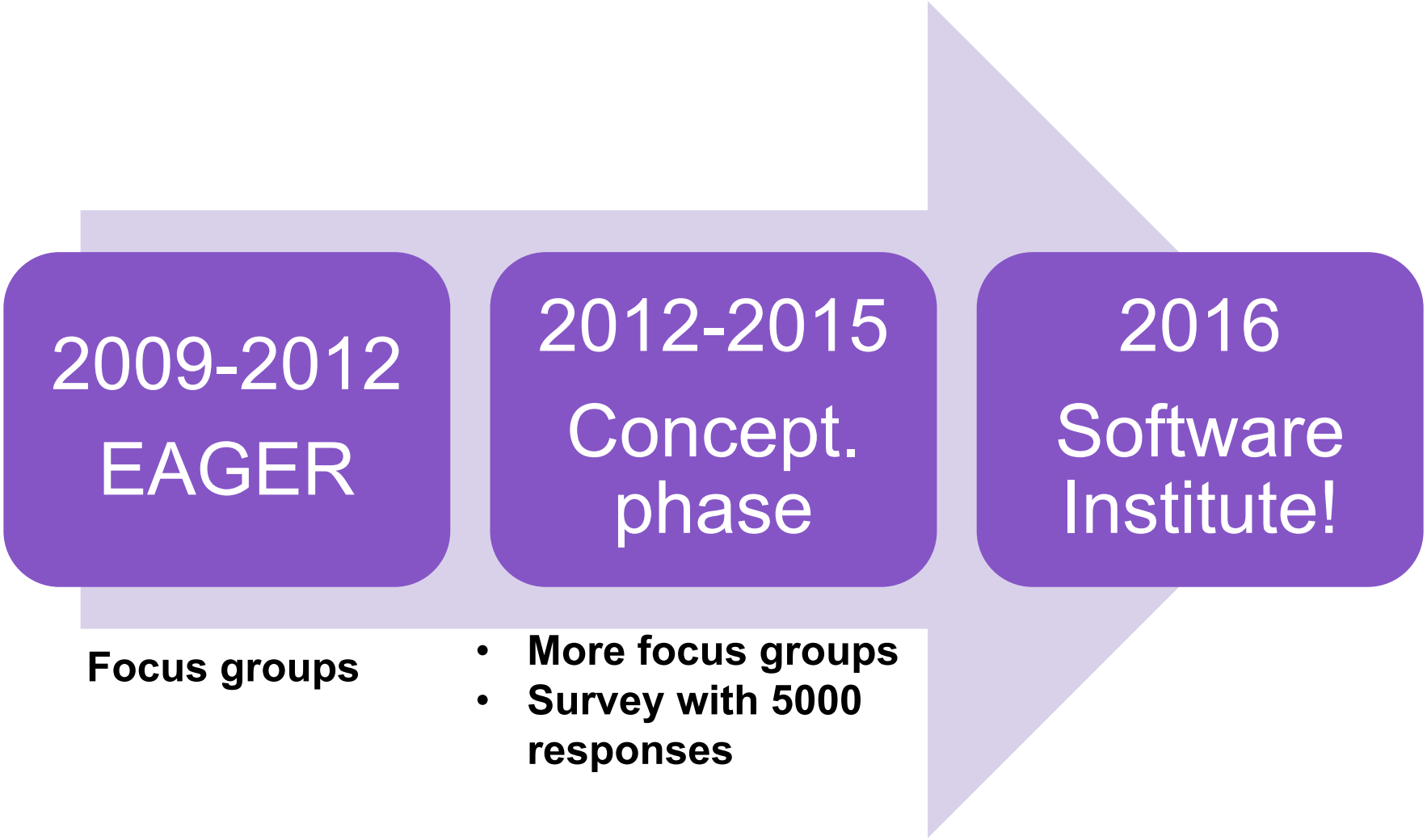
But, we observed a problem

Gateways typically developed on 3-year research funding cycle

- Developers typically
  - work in isolation
  - must bridge to variety of resources
  - need building blocks in order to focus on higher-level functionality
  - struggle to secure sustainable funding



So we studied the problem
And studied it some more



Focus group topics

1. **Characteristics of successful gateways**
   - Review of 38 NSF-funded CI reports, recommendations from NSF's CICC, scholars in the field, coverage across directorates
2. **Fields ready for transformation by gateways**
   - Snowball method of asking colleagues and experts for recommendations of both possible participants and people who could recommend others
3. **Research initiatives that have been successful and sustainable in multiple fields and through multiple funding sources**
4. **External perspectives on the evaluation criteria and compelling features of potentially successful and sustainable technology projects**
   - Suggestions of specific people and projects as well as suggestions of broader areas of expertise (e.g., non-profit marketing, open-source projects) that were made by the first two focus groups
5. **Viability of preliminary findings in a federal agency environment**
   - Program officer for recommendations within and outside of NSF, research websites of all other potentially relevant federal agencies (e.g., DOE, DOD, NOAA, EPA, etc.) to identify units (and people within those units) that were concerned with gateway-type technologies

Studied mountains of reports to select great attendees

Projects

- National Science Digital Library
- iPlant
- Earth System Grid
- TAPoR (Text Analysis Portal for Research)
- GridChem
- GISolve
- Linked Environments for Atmospheric Discovery (LEAD)
- National Snow and Ice Data Center
- nanoHUB
- VORTEX WINDS (A Virtual Organization to Reduce the Toll of Extreme WINDS on Society)
- CIPRES (Cyberinfrastructure for Phylogenetic Research)
- MyExperiment
- FLOSSmole (Free, libre, and open source software project)
- Galaxy Zoo
- Drupal, a gateway-building technology
- ScienceForCitizens.net
- Protein Data Bank (PDB)
- Long Term Ecological Research Network Office (LTER) and University of Virginia
- Folding@Home
- Sage Bionetworks
- eBird, at the Cornell University Lab of Ornithology
- National Digital Information Infrastructure and Preservation Program (NDIIPP) in the Office of Strategic Initiatives at the Library of Congress
- Dark Energy Survey
- Library of Congress's World Digital Library
- Sakai
- 18thConnect
- Computational and Information Systems Laboratory (CISL) Research Data Archive, at the National Center for Atmospheric Research
- Networked Environment for Music Analysis (NEMA)

Organizations

- Virtual Knowledge Studio for the Humanities and Social Sciences, Netherlands
- Centre for e-Science, University of Lancaster, UK
- Electronic Visualization Laboratory (EVL), University of Illinois at Chicago
- Illinois Center for Computing in Humanities, Arts, and Social Science (I-CHASS), National Center for Supercomputing Applications (NCSA)
- Cyberenvironments and Technologies group, NCSA
- Center for Public Policy, University of Houston
- Committee on the Conceptual and Historical Studies of Science, and Fellow at the Computation Institute, University of Chicago
- University of Michigan, School of Information and Provost's Office
- Mass General Institute for Neurodegenerative Disease
- CyberGIS Software Integration for Sustained Geospatial Innovation
- Common Solutions Group
- Hawkshurst Group
- Oxford e-Research Centre, UK
- Elsevier Labs
- Science House
- Open Health Tools
- Microsoft Research Connections
- Pew Internet&American Life Project
- Citizen Science Central, hosted by the Cornell Lab of Ornithology, Department of Program Development and Evaluation
- Andrew W. Mellon Foundation's Department of Scholarly Communications and Information Technology
- IMeKo
- New York Times

Agencies

- Department of Defense (multiple units)
- Department of Energy (multiple units)
- Institute of Museum and Library Services, Office of Library Services
- Library of Congress
- National Endowment for the Humanities, Office of Digital Humanities
- National Institute of Food and Agriculture (NIFA), US Department of Agriculture (multiple units)
- National Oceanic and Atmospheric Administration (NOAA), US Department of Commerce
- National Science Foundation (multiple directorates)
- United States Geological Survey, Core Science Systems

Engagement with one another

- Diverse groups of people **exchanging ideas is more effective** for eliciting novel observations and solutions to well-established challenges
- Activities designed to encourage the cross-fertilization of ideas and to move participants from concrete examples to generalizable principles

Focus group activities

- Spatial warm-ups
  - Gateway headquarters, gateway user base, length of time in existence
- Wandering flip charts with voting
  - How did you handle X that contributed to your success? What is most essential?
    - X = Tools, content, target audience, partnerships, community engagement, discipline culture, org structure
- World Café
  - Note-taking on paper table cloths, all but one person rotate to cross-fertilize
    - Sample questions: what was easy and what was hard about your project? Imagine you've been brought in as a consultant to a new gateway, how would you advise a newcomer based on your experiences? Imagine you've been brought in a sustainability adviser, how would you tell someone to navigate from development to operation, who needs to be involved?
- Structured brainstorming
  - Identify external forces that influence sustainability of gateways, opportunities and challenges
    - Funding, publication venues, evolution of scholarship, demand, technology, partnerships, education trends
- Sales pitch
  - For FG looking at new gateways, create a sales pitch for the perfect gateway
- Create a solicitation
  - For funders, if anything were feasible, what key points in a solicitation would set a gateway up for success? What changes would be necessary in a constrained environment?

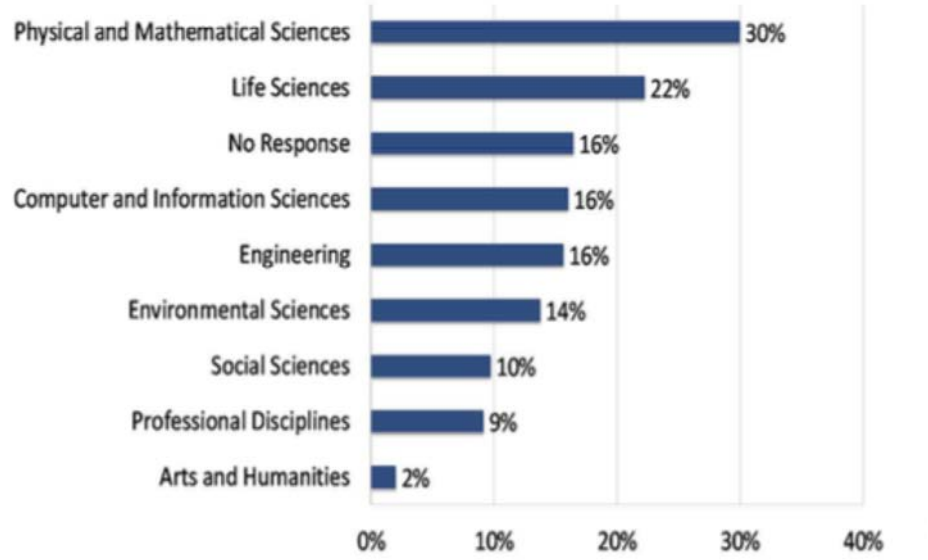
Gradually, pictures emerged

Next we wanted to test our hypotheses on a larger scale

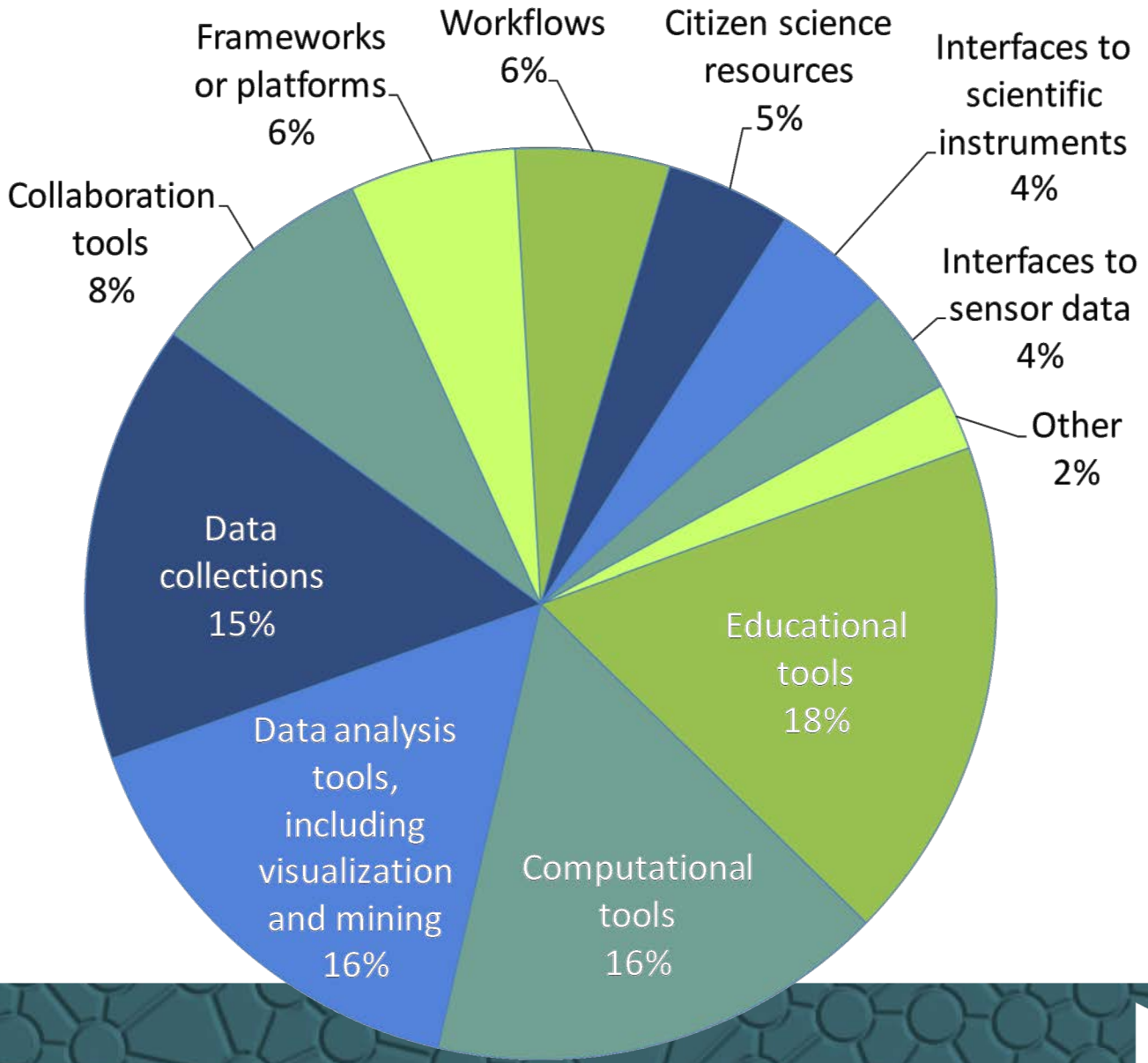
SGCI

Large-scale survey launched in 2014
Sent to 29k NSF PIs and academic CIOs and CTOs

- 5000 responses!
- 58 domain areas across 9 broader categories
- Who's using gateways?
- For what?
- Who's developing gateways?
- What do they need?

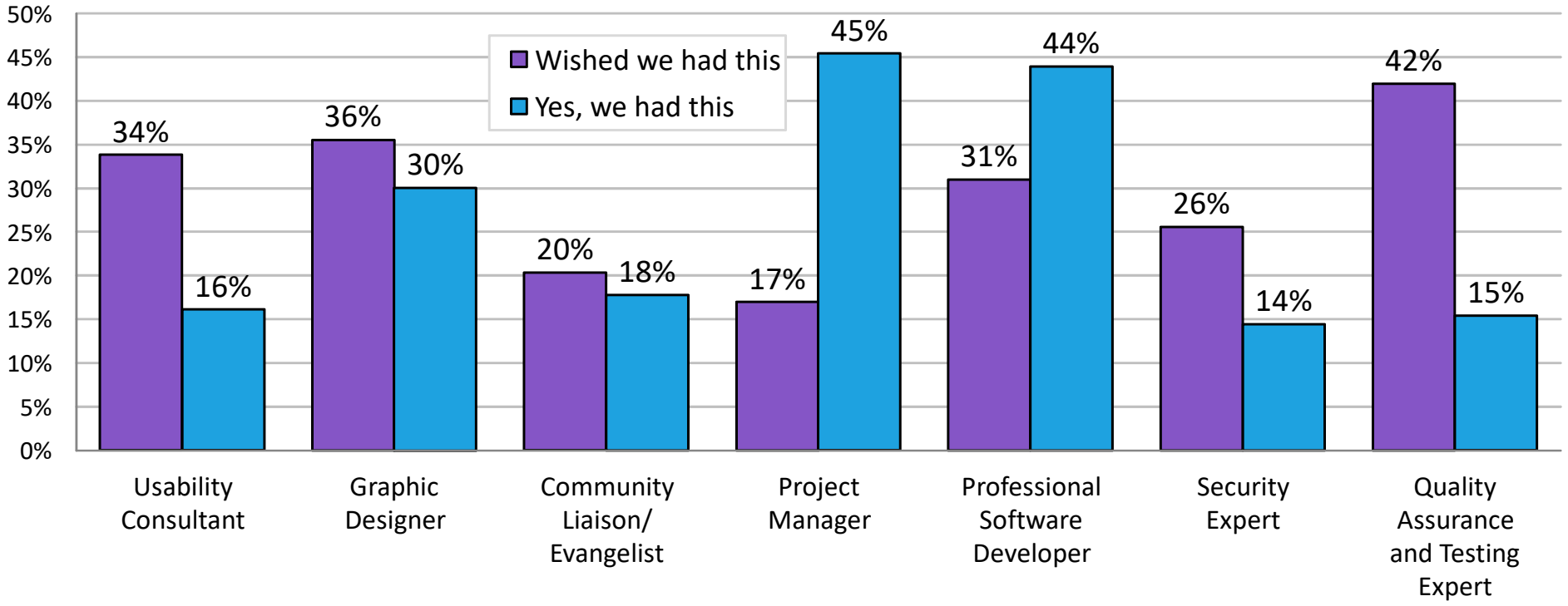


Gateways built for many purposes
Results of 5000-person survey in 2014



Why is it difficult to build a science gateway?

- Building a gateway takes many types of expertise
  - But projects cannot always locate or afford to hire these specialists
  - Finding the right people for short-term work is also difficult
- Need for a variety of team members, like a start-up company!

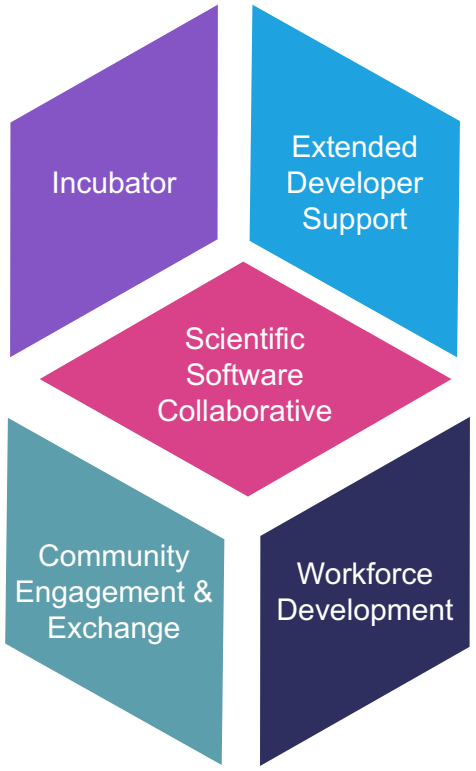


Access to specialized services can help

- Need for a variety of team members, like a start-up company!
  - Finding the right people for short-term work is difficult
- Many topics well-suited for short-term consulting

Service	% Interest
Evaluation, impact analysis, website analytics	72%
Adapting technologies	67%
Web/visual/graphic design	67%
Choosing technologies	66%
Usability Services	66%
Developing open-source software	64%
Support for education	64%
Keeping your project running	62%
Legal perspectives	61%
Managing data	60%
Cybersecurity consultation	57%
Website construction	57%
Software engineering process consultation	53%
Source code review and/or audit	51%
High band-width networks	45%
Scientific instruments or data streams	44%
Management aspects of a project	38%

A vision for an institute emerged
A holistic, service-based institute



- Diverse skills, on demand
- Hands-on help building gateways
- An app store for gateways!
- Interactions with others building gateways
  - Gateways conference
  - Special journal issue with int'l partners
- Opportunities for students

A unique platform-independent approach to gateway development.
We recommend what's best for the client.

Planning has paid off

Activities at 18-months

575 webinar attendees & registrants

220 conference attendees

180 student &

21 faculty participants

65 letters of collaboration

38 Bootcamp attendees

86 tutorials & outreach presentations

18 Incubator consultations

8 partners

20 Extended Developer Support projects

10 Campus Gateway Group engagements

402 Gateway Catalog entries

SGCI

Early gateway-building clients

Software as a Service

COSMIC2
Michael Cianfrocco, University of Michigan

LSU Systems Biology
Michal Brylinski, Louisiana State University

SimCCS
Kevin Ellet, Indiana Geological and Water Survey

ChemCompute
Mark Perri, Sonoma State University

nSides
Rami Vanguri, Columbia University

Interactive Parallelization Tool
Ritu Arora, University of Texas

Cyberinfrastructure

Galaxy CloudLaunch
Enis Afgan, Johns Hopkins University

Data Distribution

Coastal Emergency Risk Assessment
Carola Keiser, LSU and Jason Fleming, SCS

CitSci.org
Greg Newman, Colorado State University

ENIGMA
Lisa Eyler, University of California, San Diego

Ocean Observatories Initiative
Ivan Rodero, Rutgers University

Aquavit
Jack Smith, Marshall University

Collaboration

QUBES
Drew LaMar, William& Mary

Ecology Plus
Teresa Mourad, Ecological Society of America

SGCI

Opportunities to Engage

- Request services
  - Short term consulting, longer term gateway-building
- Find or list a gateway
  - catalog.sciencegateways.org
- Request a Letter of Commitment to leverage existing SGCI offerings in proposal
- Become involved as a partner or affiliate
  - Engaging with SGCI clients
- Train students via the internship program

www.sciencegateways.org

URSSI observations very relevant to gateway developers

- ✓Research software is essential to progress in almost all fields
- ✓Often not developed in an efficient or sustainable way
  - ✓Knowledge locked away in individual laboratories or shared via method papers
- ✓Developers of research software often don't use best practices that ease development, maintainability, sustainability, reproducibility
- ✓Developers don't match the diversity of overall society or of user communities

Gateways are examples of research software

- Many of the concepts that this group advances will help the gateway community
- We are also working to support career paths
- Gateways are often not developed efficiently
  - Siloed activities
  - Not leveraging diverse expertise
- We feel campus “mini institutes” may be one solution
  - Ideal connection via Sandra Gesing
    - Leads SGCI campus program
    - URSSI co-PI
- Student programs
  - **Skills pipeline**
  - Campus programs will provide broad paths for our students
- I believe our activities could really complement one another

Gateway Skills Development Study

Non- Technical Skills

Internships, mentoring roles, apprenticeships, etc

Adopt open source practices

Communication Skills

Code of Research Ethics

Basic Technical Skills

Software

- Programming languages including Python, Java, • Javascript, Ruby, PHP and C.
- Multiple Web frameworks such as Django.

Basic Software Engineering

- Using a version control system like Git effectively
- Having a reliable, reproducible way of compiling or building your software.
- Having tests for your software

System Administration

- LINUX or UNIX prompt and basic commands
- scripts in bash or Perl.
- managing accounts, installing software/Web servers,
- managing security settings such as firewalls.

SGCI

Advanced Technical Skills

Cluster and Cloud Computing	*Security in Distributed Systems* *Usability*	*Messaging Systems* *Configuration Management*
Continuous Integration and Deployment (CI/CD)	*Application Programming Interfaces (APIs) and Distributed Component Architectures*	*Distributed Systems and others*

CSCI-B 649 SCIENCE GATEWAY ARCHITECTURES

Course

Fall 2017 - Science Gateway Architectures

- **Course:** CSCI-B 649, Topics in Systems, Computer Science, School of Informatics and Computing, Indiana University
- **Instructors:** Marlon Pierce, marpierc@iu.edu; Suresh Marru, smarru@iu.edu
- **Associate Instructors:** Sneha Tilak; Eldho Mathulla
- **Class Schedule** Tuesdays and Thursdays from 4 pm to 5.15pm in I2 (Informatics East) Room 150
- **Office Hours** On Request
- **HipChat Instant Messaging** https://www.hipchat.com/g6Use7j8w

URSSI discussion topics also relevant for gateway developers

- Usability
- Importance of software/gateways in research
- Diversity
- Faculty contributions
- Sustainability models
- Best practices
- Grant funding

Thank you

- I hope our experiences will be helpful for URSSI
- I'm excited to be involved in the process to design an institute of value
- I'm looking forward to the discussions throughout the week

www.sciencegateways.org