

What are the benefits of using Science Gateways

Amit Majumdar, Division Director
Data Enabled Scientific Computing Division

San Diego Supercomputer Center
University of California San Diego

UCLA Workshop 2017

HPC Resources

- **Available via XSEDE – Extreme Science and Engineering Discovery Environment**
 - NSF funded supercomputers, advanced support, services, allocation, EOT
- **You know or will learn – how to**
 - Compile codes
 - Launch and manage jobs
 - Manage data on filesystems
 - Use HPC resources effectively
 - etc

Administrative and Technical tasks (barriers?)

- **Write allocation proposals (peer-reviewed) for supercomputer time every year**
- **Understand HPC machines, policies, complex OS/software**
- **Install and benchmark complex applications on HPC resources**
- **Different machines have different schedulers**
- **Understand and manage remote authentication**
- **Figure out data transfer, file systems, storage**



Science Gateways

A Science Gateway is a community-developed set of tools, applications, and data that are integrated via a portal or a suite of applications, usually in a graphical user interface, that is further customized to meet the needs of a specific community. Gateways enable entire communities of users associated with a common discipline to use national resources through a common interface that is configured for optimal use. Researchers can focus on their scientific goals and less on assembling the cyberinfrastructure they require. Gateways can also foster collaborations and the exchange of ideas among researchers.

How to Turn Your Project into a Science Gateway

1. Get an XSEDE allocation; Start-up and Educational allocations require only a one paragraph project description. For more information visit the [Allocations section](#) of the web site.
2. [Register your project as an XSEDE Gateway](#)
3. [Build a portal](#)
4. Set up your developer accounts by [Adding users to an existing allocation](#). Also, set up your [Gateway community accounts](#).

Using Existing Gateways

Gateways are independent projects, each with its own guidelines for access. Most gateways are available for use by anyone, although they usually target a particular research audience. XSEDE Science Gateways are portals to computational and data services and resources across a wide range of science domains for researchers, engineers, educators, and students. Depending on the needs of the communities, a gateway may provide any of the following features:

Science Gateways

[Overview](#)[Gateway Listing](#)[For Pls](#)[For Developers](#)[Gateways Symposium](#)[Gateway Applications](#)

A Science Gateway

1. **Easy user interface to HPC for a domain science**
 1. Primarily secure online web portal
 2. Programmatically also in some cases
2. **Easy web click and submit mode to upload model/input data, run of codes on HPC, notification**
3. **User support from Science Gateway staff**
4. **Provide widely used domain science tools on the backend on HPC resources - optimally installed**
5. **Ability to easily get to the results, download results**
6. ***No writing of allocation proposals (SGW PI does that for the community)***

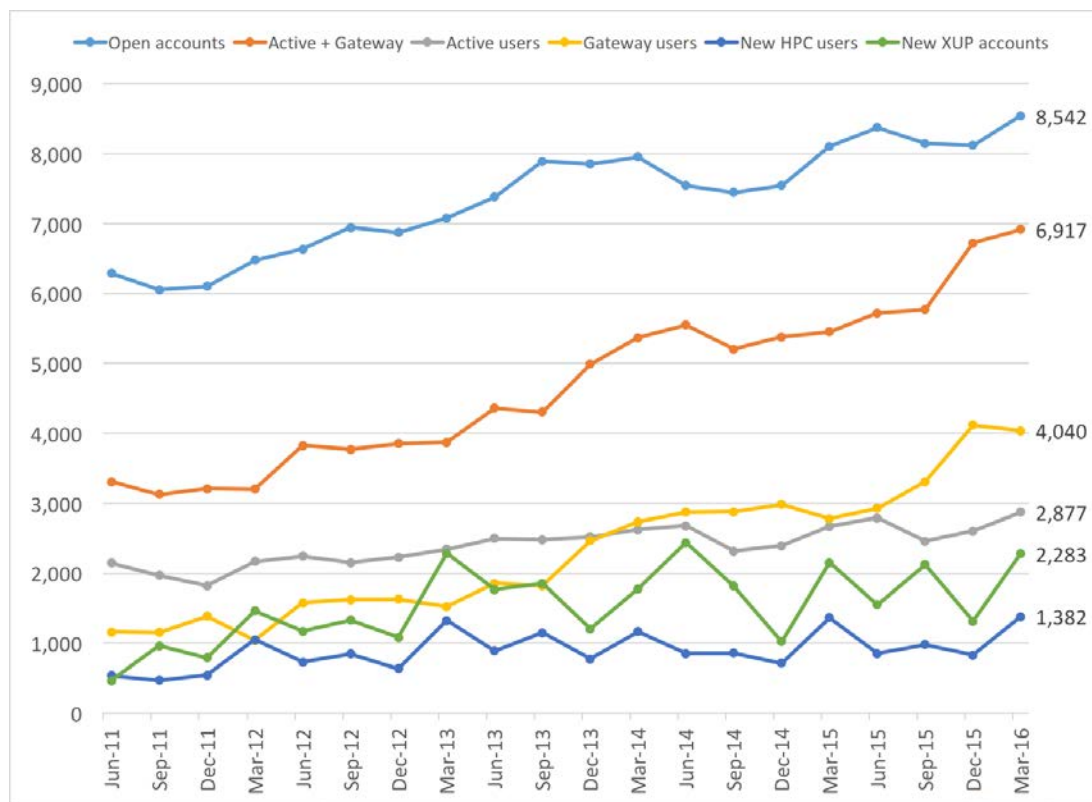
A Science Gateway catalyzes and democratizes research for everybody including researchers and students from institutions with less resources

TITLE	FIELD OF SCIENCE	PORTAL HOMEPAGE
Diagrid	Advanced Scientific Computing	Visit Portal
UCI Complex Social Science Gateway	Anthropology	Visit Portal
Massive Pulsar Surveys using the Arecibo L-band Feed Array (ALFA)	Astronomical Sciences	Visit Portal
Center for Multiscale Modeling of Atmospheric Processes	Atmospheric Sciences	Visit Portal
Biodrugscore: A portal for customized scoring and ranking of molecules docked to the human proteome	Biochemistry and Molecular Structure and Function	Visit Portal
High-Resolution Modeling of Hydrodynamic Experiments with UltraScan	Biophysics	Visit Portal
ROSIE, The Rosetta Online Server that Includes Everyone	Biophysics	Visit Portal
Chem Compute	Chemistry	Visit Portal
Computational Chemistry Grid (GridChem)	Chemistry	Visit Portal
ParamChem Gateway	Chemistry	Visit Portal
User-Friendly Security Solutions for Grid Environments	Communications and Computational Systems	Visit Portal
OpenTopography	Earth Sciences	Visit Portal
Purdue Environmental Data Portal	Earth Sciences	Visit Portal
WaterHUB - Platform for water education, research, data access, partnership and collaboration	Earth Sciences	Visit Portal
Network for Earthquake Engineering Simulation	Earthquake Hazard Mitigation	Visit Portal
CMS gWMS	Elementary Particle Physics	Visit Portal
Network for Computational Nanotechnology and nanoHUB	Emerging Technologies Initiation	Visit Portal
Globus Online	Engineering Infrastructure Development	Visit Portal
CyberGIS Gateway	Geography and Regional Science	Visit Portal
The Earth System Grid	Global Atmospheric Research	Visit Portal

Network for Computational Nanotechnology and nanoHUB	Emerging Technologies Initiation	Visit Portal
Globus Online	Engineering Infrastructure Development	Visit Portal
CyberGIS Gateway	Geography and Regional Science	Visit Portal
The Earth System Grid	Global Atmospheric Research	Visit Portal
The iPlant Collaborative Agave API	Integrative Biology and Neuroscience	Visit Portal
MP-Complete	Materials Research	Visit Portal
VLab - Virtual Laboratory for Earth and Planetary Materials	Materials Research	Visit Portal
NIST Digital Repository of Mathematical Formulae	Mathematical Sciences	Visit Portal
Galaxy	Molecular Biosciences	Visit Portal
Integrated database and search engine for systems biology (IntegromeDB)	Molecular Biosciences	Visit Portal
ROBETTA: Automated Prediction of Protein Structure and Interactions	Molecular Biosciences	Visit Portal
Providing a Neuroscience Gateway	Neuroscience Biology	Visit Portal
SCEC Earthworks Project	Seismology	Visit Portal
Astero seismic Modeling Portal	Stellar Astronomy and Astrophysics	Visit Portal
CIPRES Portal for inference of large phylogenetic trees	Systematic and Population Biology	Visit Portal
Computational Anatomy	Visualization, Graphics, and Image Processing	Visit Portal

Gateways are revolutionizing and democratizing supercomputer access

- Users can access supercomputers via a browser
- Gateway users exceed login users on XSEDE resources



← Sum

← Gateway

← Login

NSF Awards \$15 Million to Create Science Gateways Community Institute

The Institute will accelerate the development and application of highly functional, sustainable science gateways that address the needs of researchers across the full spectrum of NSF directorates

READ MORE



Incubator

Science Gateways Community Institute

a synergistic focal point



Innovate, Educate, Collaborate:

FOR
UC/UCSD Researchers

FOR
National HPC Users

FOR
Industry & Sponsors

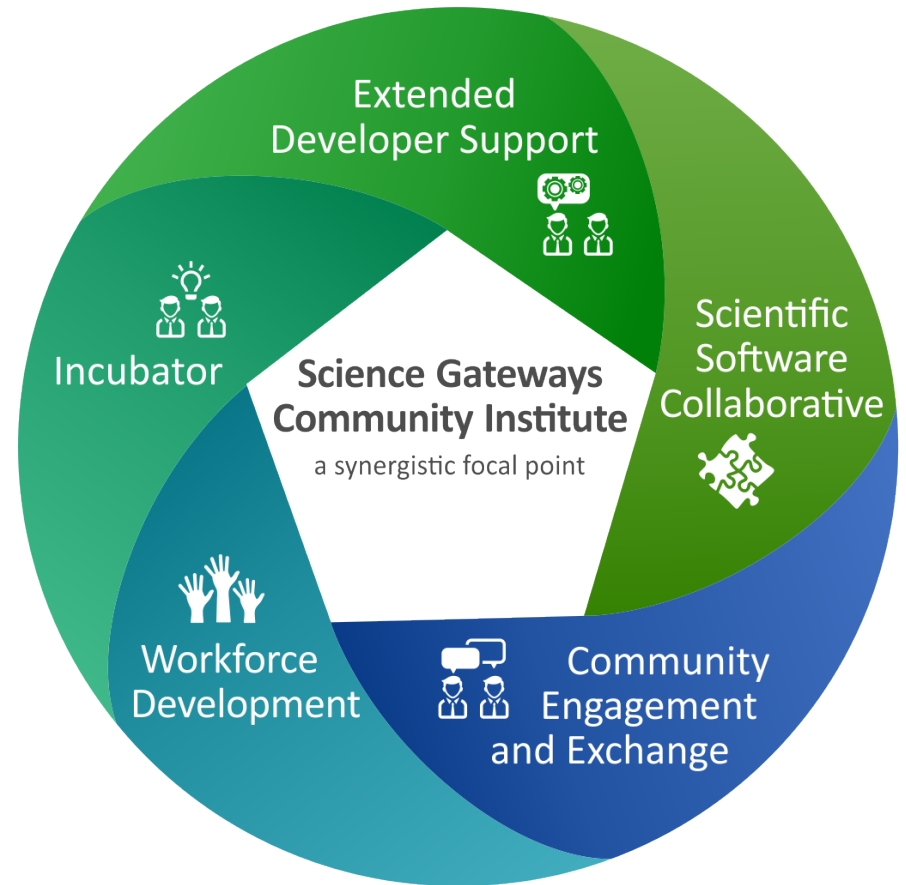
FOR
Students & Educators

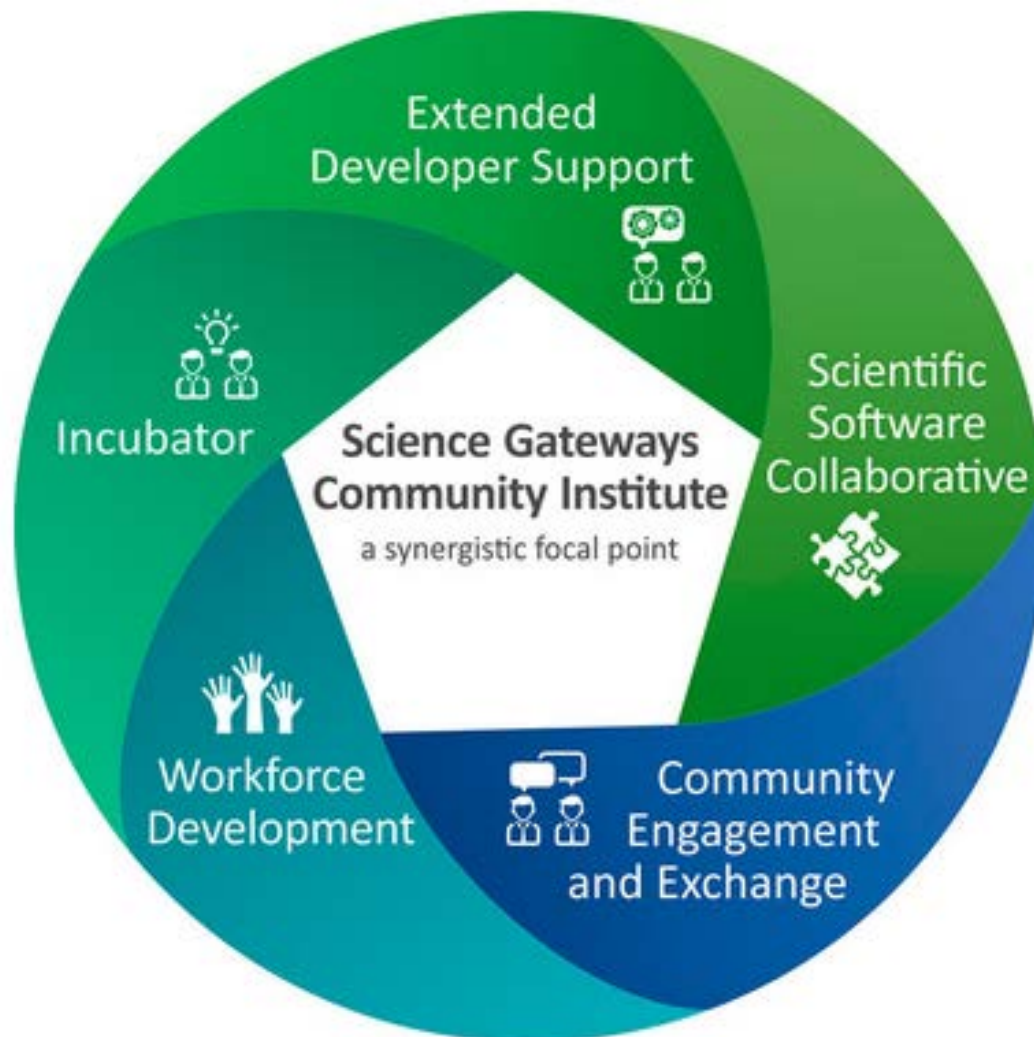
Nancy Wilkins-Diehr, SDSC – PI - <http://sciencegateways.org/>

Other institutions: Elizabeth City State in North Carolina, Indiana University, University of Notre Dame, Purdue University, the Texas Advanced Computing Center at the University of Texas, Austin, and the University of Michigan at Ann Arbor

SDSC is taking gateway leadership to the next level with the Science Gateways Community Institute

- New \$15M NSF award for 5 years with 5-year renewal option
- Led by PI Nancy Wilkins-Diehr of SDSC
- Co-PIs: Maytal Dahan (TACC), Linda Hayden (ECSU), Katherine Lawrence (Michigan), Marlon Pierce (Indiana), Michael Zentner (Purdue)
- Goes beyond ECSS to support
 - Overall gateway development, both front end and back end
 - Gateways that do not use HPC

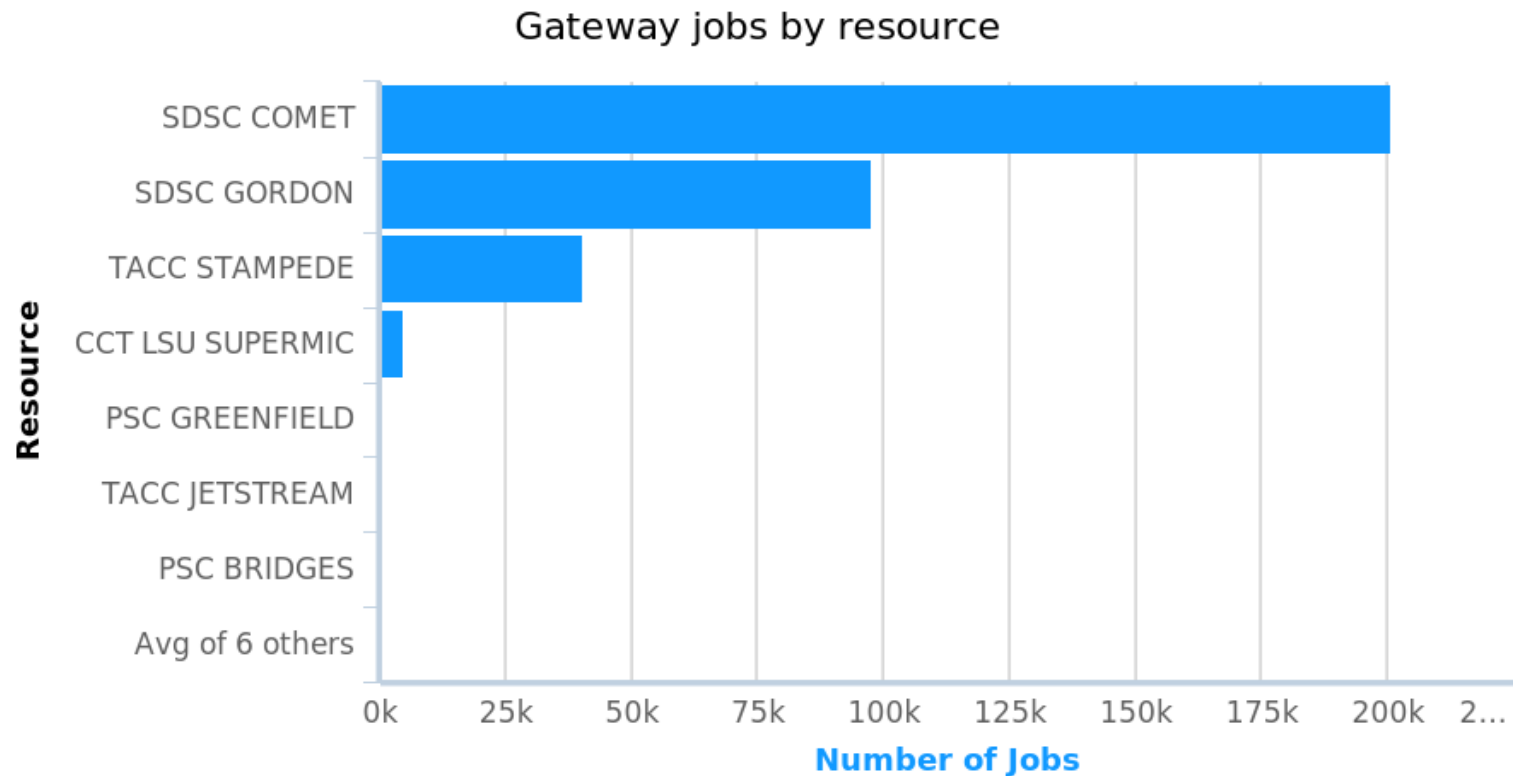




The five key areas for the Science Gateways Community Institute to increase the number, ease of use, and effective application of gateways to serve the greater research and engineering community. *Source: SDSC*

SDSC leads in hosting gateways

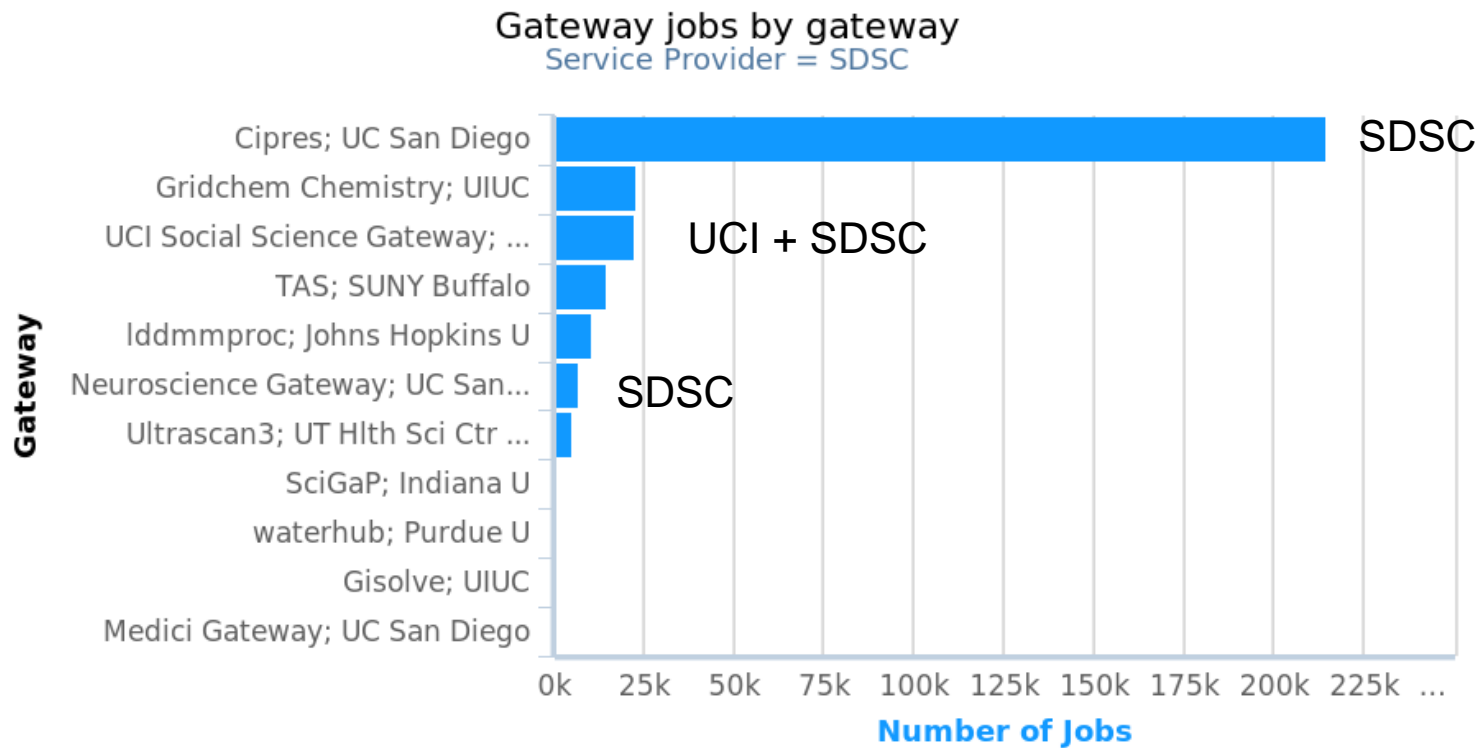
- Comet and Gordon accounted for 86% of gateway jobs on XSEDE resources over the past year



2015-09-01 to 2016-09-30 Src: XDCDB. Powered by XDMoD/Highcharts

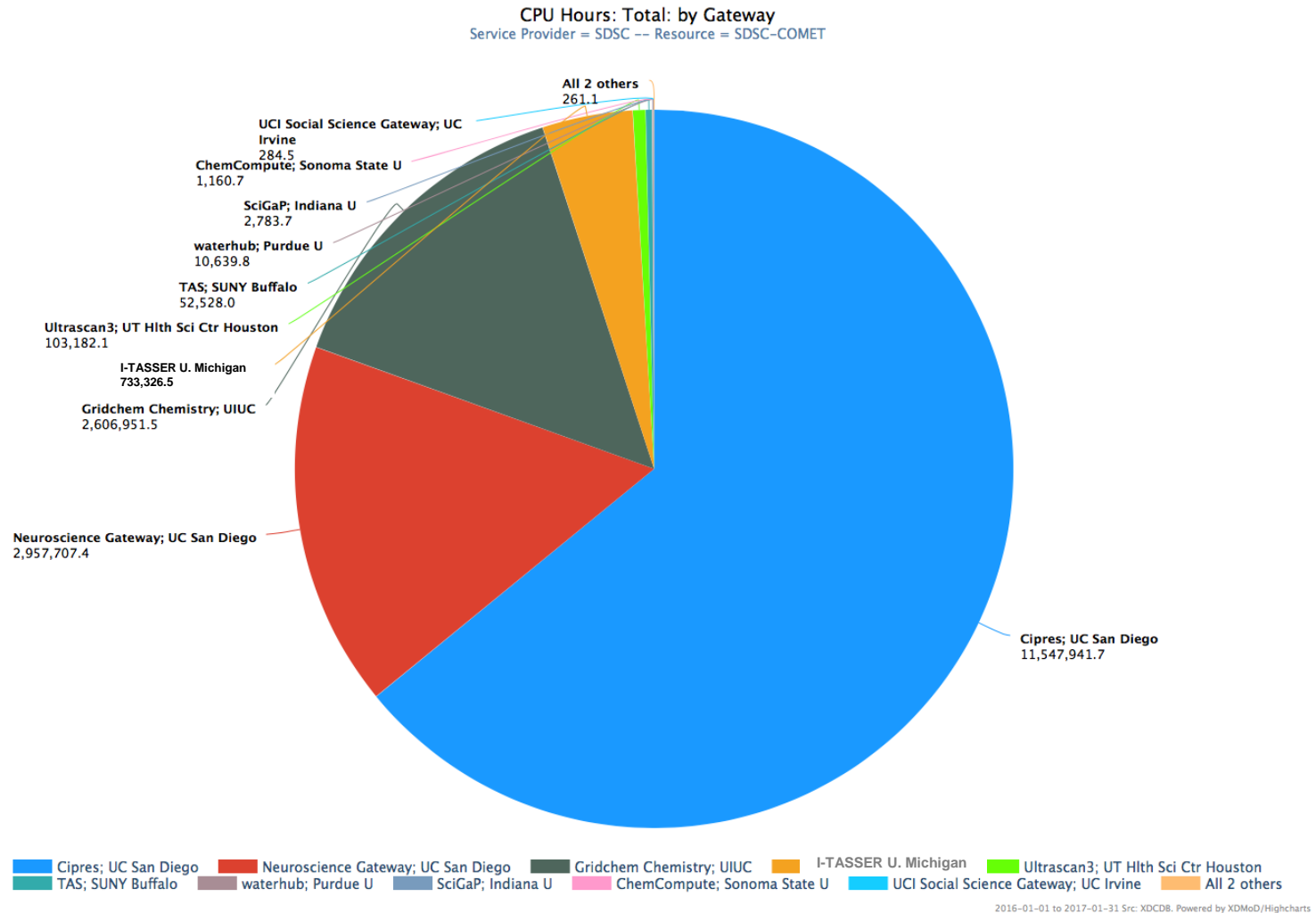
SDSC leads in developing and maintaining gateways

- 3 of top 7 gateways that SDSC hosts were developed at or together with SDSC



2015-09-01 to 2016-09-30 Src: XDCDB. Powered by XDMoD/Highcharts

There are now 12 gateways on Comet across several domains. More are coming.



The advent of DNA sequencing lets scientists infer phylogenetic trees from multiple sequence alignments

Multiple sequence alignment is a matrix of taxa vs characters

Human	AAGCTTCACCGGCGCAGTCATTCTCATAAT...
Chimpanzee	AAGCTTCACCGGCGCAATATCCTCATAAT...
Gorilla	AAGCTTCACCGGCGCAGTTGTTCTTATAAT...
Orangutan	AAGCTTCACCGGCGCAACCACCCTCATGAT...
Gibbon	AAGCTTTACAGGTGCAACCGTCCTCATAAT...



Tree inference code, e.g., BEAST, MrBayes, or RAxML



Phylogeny is represented as tree with taxa at its tips



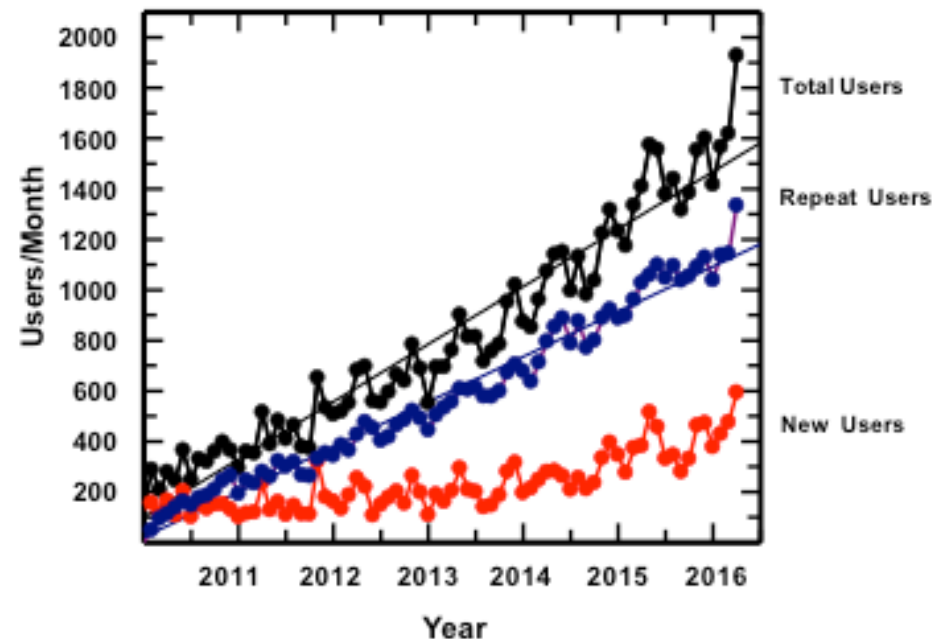
SDSC develops, maintains, & hosts the NSF-funded CIPRES gateway, which runs phylogenetics codes via a browser interface; analyses span the entire tree of life

Mark Miller, Wayne Pfeiffer, Terri Schwartz, SDSC

The screenshot shows a web browser window with the address bar displaying `https://www.phylo.org/portal2/createTask/create.action`. The page title is "Cipres Science Gateway | Home". The browser's address bar includes a "Reader" button and a search icon. Below the address bar is a navigation bar with links: "At Your Servi...ne : Sign In", "Apple", ".Mac", "Amazon", "News", "eBay", "Yahoo!", "Google Maps", "YouTube", "Wikipedia", and "Popular". The main header features the CIPRES logo (a tree icon) and the text "CIPRES SCIENCE GATEWAY". To the right of the logo are logos for "iPlant Collaborative" and "XSEDE Extreme Science and Engineering Discovery Environment". Below the header is a dark blue navigation bar with links: "CIPRES", "Home", "Toolkit", "My Profile", "Help", "How to Cite Us", "XSEDE Status", and "Logout". The main content area is divided into two columns. The left column, titled "Folders", shows "Total Storage: 67 MB" and a list of folders: "FastTree", "BEAST", "nexus-to-phyli", "MrBayes", "Kurt", "ReadSeq", "RAxML", "Data (2)", and "Tasks (7)". The right column, titled "Create new task", has four tabs: "Task Summary" (selected), "Select Data", "Select Tool", and "Set Parameters". Below the tabs is a message: "You may edit your task using the tabs above. Current CPU Hr Usage: 45 [Explain this?](#)". Below this is a form with the following fields and buttons: "Description" (text input), "Input" (button "Select Input Data"), "Tool" (button "Select Tool"), "Input Parameters" (button "Set Parameters"), and three buttons at the bottom: "Save Task" (green), "Save and Run Task" (green), and "Discard Task" (blue). At the bottom of the form is a note: "Saved tasks can be run later from the task list. XSEDE tasks are limited to 168 hours. Non-XSEDE tasks are limited to 72 hours."

The CIPRES gateway has been extremely popular and supports thousands of researchers around the world

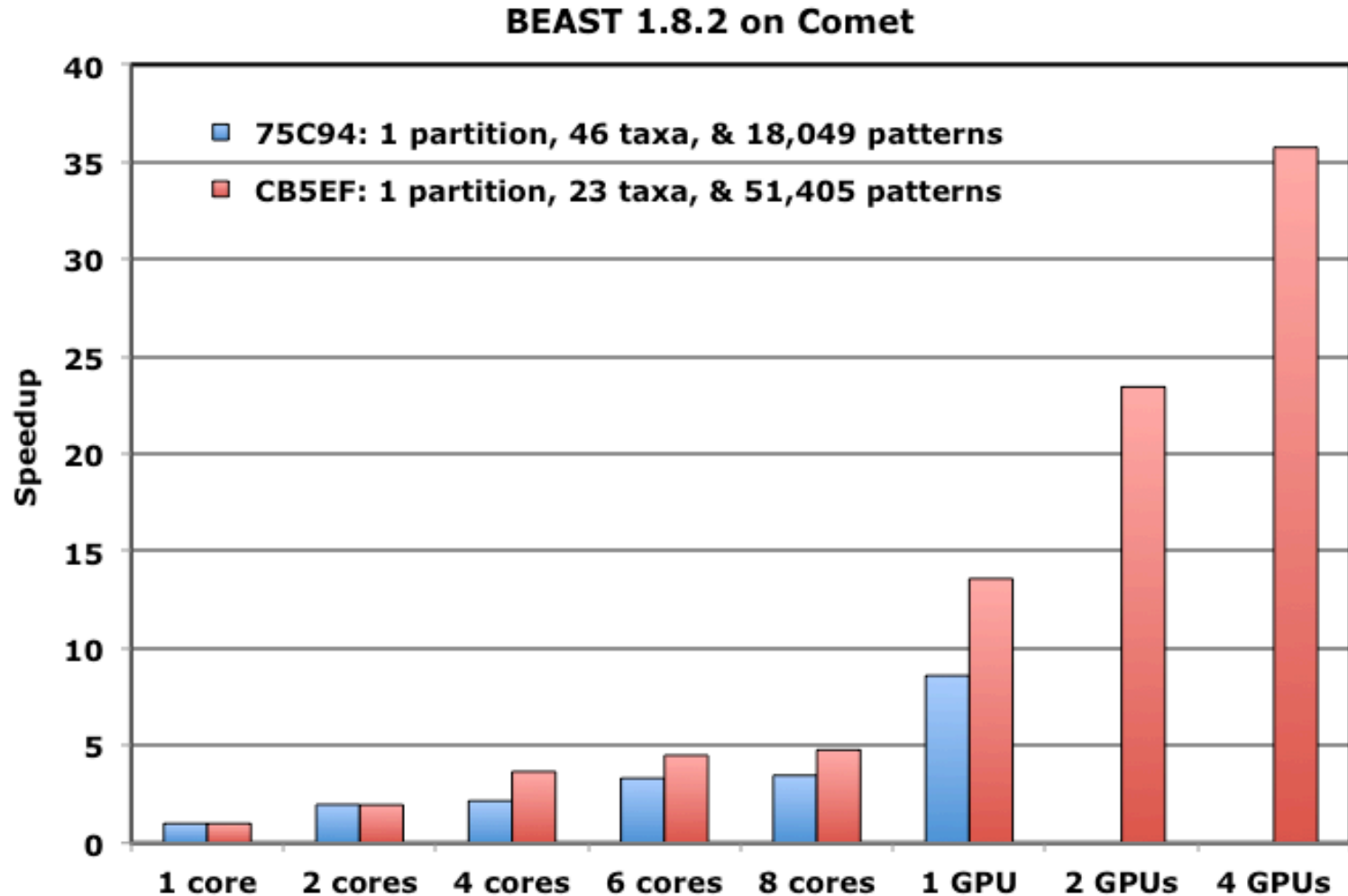
- >17,000 CIPRES users have run on NSF-funded supercomputers, including 3,152 in 4Q2015 or 47% of all active XSEDE users!
- >2,000 publications have been enabled by CIPRES use!
- **US statistics from 2015**
 - 49 states + 2 territories + DC
 - 252 universities & colleges
 - 18 institutes
 - 22 museums, gardens, & zoos
 - 21 government agencies
 - 4 high schools
- **Non-US statistics from 2015**
 - 85 countries
 - 603 universities & colleges
 - 161 institutes
 - 80 museums, gardens, & zoos
 - 134 government agencies



11 codes are supported by CIPRES on Comet & Gordon; most have modest scalability; some run for days

Code	Version	Language	Computer	Cores charged
BEAST	1.8.3	Java + C++	Comet	2, 4, or GPUs
BEAST2	2.3.2	Java + C++	Comet	1, 2, or 3
DPPDiv	1.0	C++	Gordon	16
FastTree	2.1.8	C	Comet	3
GARLI	2.0.1	C++	Comet	1 to 24
jModelTest2	2.1.6	Java + C	Gordon	8
MAFFT	7.187	C	Gordon	8
MrBayes	3.2.6	C + C++	Gordon	8 or 16
Migrate	3.6.11	C	Comet	1 to 72
Phylobayes	1.5a	C++	Gordon	64
RAxML	8.2.8	C	Comet	12, 24, or 48

BEAST phylogenetic tree inference for data sets with many patterns speeds up very well using K80 GPUs instead of Haswell cores on Comet



NSF funded Neuroscience Gateway at SDSC

Amit Majumdar, Subhashini Sivagnanam, Kenneth Yoshimoto, SDSC

- **NSG – in operation since early 2013 – nsgportal.org**
- **Built using the CIPRES gateway software**
- **NSG benefits the broader neuroscience research community in several ways, e.g.:**
- **Researchers can run larger complex neuronal networks, parameter sweep simulations, brain image processing tools**
- **Fully integrated The Virtual Brain (TVB) connectome pipeline workloads can be processed in parallel**
- **Easy access to widely used simulation tools such as:
Brian, NEST, NEURON, pGENESIS, PyNN, MOOSE, FreeSurfer, TensorFlow, Octave, CARLSim, R, Matlab (coming in the future)**
- **Researchers from EU Human Brain Project providing optimization pipeline**
- **Access to new HPC resources – GPUs, Intel MICs**
- **Can be used by researchers with limited local (university-level) resources to address questions that require access to large scale, advanced systems**
- **Can be used by simulator developers to test, benchmark, and scale codes on large scale resources**
- **Can be used for classes, workshops, and tutorials**

[BluePyOpt on Comet](#) (1.1.27) ⓘ - Running BluePyOpt analyses

[Brian on Stampede](#) (2.0b2) ⓘ - Brian is a simulator for spiking neural networks

[Brian on Comet](#) (2.0b2) ⓘ - Brian is a simulator for spiking neural networks

[The Virtual Brain Personalized Multimodal Connectome Pipeline on Comet](#) () ⓘ - Connectome Pipeline on Comet

[FREESURFER on Comet](#) (5.3.0) ⓘ - Freesurfer tool on Comet

[PyMOOSE](#) (3.0.1 Gulab Jamun) ⓘ - Running Moose models on Comet

[NEST on Stampede](#) (2.6.0) ⓘ - Neural Simulation Technology using Python

[NEST using Python on Comet](#) (2.2.1) ⓘ - Neural Simulation Technology using Python

[NEST on Stampede](#) (2.6.0) ⓘ - Neural Simulation Technology

[NEST on Comet](#) (2.2.1) ⓘ - Neural Simulation Technology

[NEURON7.3 Python on Stampede](#) (7.3) ⓘ - Using Python to run NEURON 7.3

[NEURON7.3 Python on Comet](#) (7.3) ⓘ - Using Python to run NEURON 7.3

[NEURON7.3 on Stampede](#) (7.3) ⓘ - Latest NEURON simulation software package on Stampede

[NEURON7.3 on Comet](#) (7.3) ⓘ - Latest NEURON simulation software package on Comet

[NEURON7.4 Python on Comet](#) (7.4) ⓘ - Using Python to run NEURON 7.4

[NEURON7.4 on Comet](#) (7.4) ⓘ - Latest NEURON simulation software package on Comet

[PGENESIS on Stampede](#) (2.3) ⓘ - Parallel Genesis software

[PGENESIS on Comet](#) (2.3) ⓘ - Parallel Genesis software

[PyNN on Stampede](#) (0.7.5) ⓘ - Python package for simulator-independent specification of neuronal network models

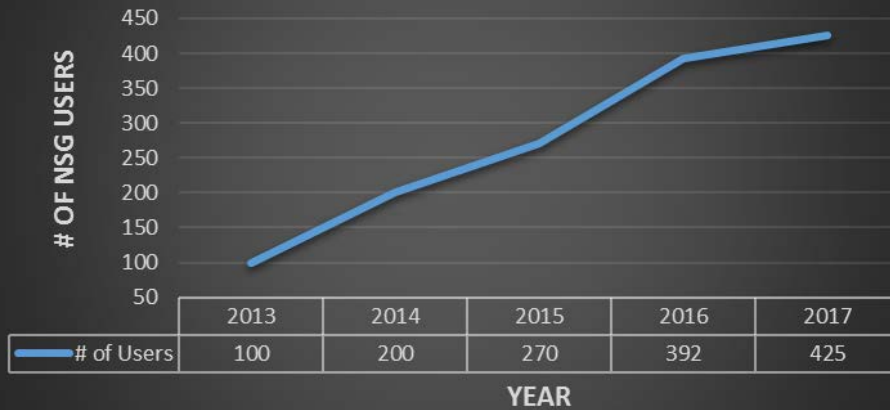
[PyNN on Comet](#) (0.7.5) ⓘ - Python package for simulator-independent specification of neuronal network models

[Python on Stampede](#) (2.7.9) ⓘ - Running Python models

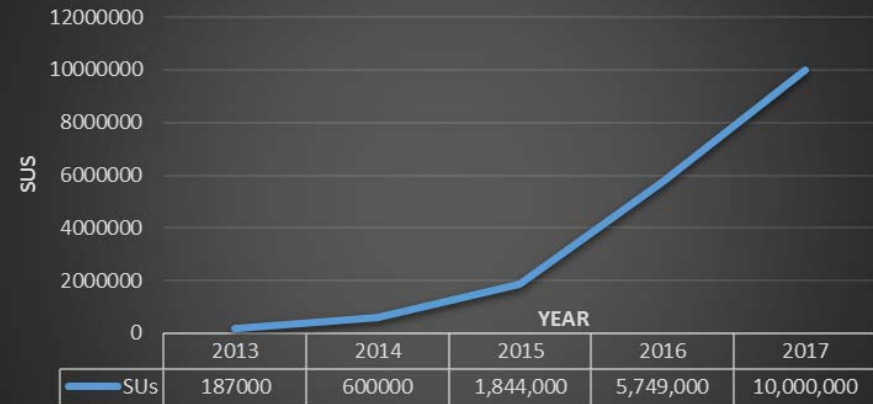
[Python on Comet](#) (2.7.9) ⓘ - Running Python models

NSG usage growing – since 2013

NSG USERS



Yearly SU usage/allocated



Maximum Core Count Job per Month (linear forecast trend curve -----)



Science Gateways - Summary

- **Allows anyone from anywhere to easily access and use HPC (and data, instrument etc.) for computational science**
 - All users start out with some amount of core hours (depends on the gateway and the science)
 - If you graduate out of a SGW, you can write your own allocation proposal (gateway/XSEDE staff can help)
 - In many cases still use the gateway to charge to your allocation
- **It creates a cyberinfrastructure environment for the science community to enable**
 - Research
 - Education
 - Sharing of information and data