

---

# ***XSEDE Overview***

**Robert Sinkovits**  
**San Diego Supercomputer Center**



*2013 Summer Institute: Discover Big Data, August 5-9, San Diego, California*

---

SAN DIEGO SUPERCOMPUTER CENTER *at the* UNIVERSITY OF CALIFORNIA, SAN DIEGO



---

## *What is XSEDE*

XSEDE (Extreme Science and Engineering Discovery Environment) is a five-year, \$121-million project supported by the National Science Foundation. The goal is to provide a single virtual system that scientists can use to interactively share computing resources, data, and expertise.

Although the XSEDE name explicitly says “Science and Engineering”, it serves a broader community that includes the social sciences, humanities and other fields

# XSEDE

Extreme Science and Engineering  
Discovery Environment

**SDSC**

*2013 Summer Institute: Discover Big Data, August 5-9, San Diego, California*

---

SAN DIEGO SUPERCOMPUTER CENTER *at the* UNIVERSITY OF CALIFORNIA, SAN DIEGO



---

## *What Does XSEDE do?*

- Centralizes many of the functions that are common to all of the participating supercomputer centers (more generally service providers)
  - Allocations, accounting & peer review of proposals
  - Help desk / ticketing system
  - Documentation
  - Web site, portal, security and authentication
- Advanced support through ECSS program
- Coordinate networking and shared XSEDE wide file system
- Education, outreach & training
- Annual meeting
- Gateways and middle ware
- Campus champions – representatives to campuses

---

## *How SDSC fits into XSEDE*

- SDSC is not a division of XSEDE nor does it work for XSEDE. Although XSEDE is an important funding source for SDSC, it is by no means the largest or dominant.
- SDSC and the other centers (TACC, PSC, NICS, etc.) are autonomous organizations that compete for supercomputing resources. Typically this is through the NSF Track 2 program, with the condition that most of cycles will be made available through the XSEDE allocations process.
- SDSC is an Organized Research Unit (ORU) at UCSD and is involved in many independent projects that are distinct and unrelated to XSEDE.

---

## Who is XSEDE?

Total of 17 organizations that provide compute, storage and other resources along with training, software and expertise



2013 Summer Institute: Discover Big Data, August 5-9, San Diego, California

---

SAN DIEGO SUPERCOMPUTER CENTER at the UNIVERSITY OF CALIFORNIA, SAN DIEGO



## ***XSEDE resources – what's available***

High performance computing resources ranging from 100-6000 TFlop peak



High throughput computing



Open Science Grid



Visualization



Distributed  
test beds

4-11 PB parallel file systems, XSEDE Wide File System, 60-170 PB tape archives



**SDSC**

2013 Summer Institute: Discover Big Data, August 5-9, San Diego, California

SAN DIEGO SUPERCOMPUTER CENTER at the UNIVERSITY OF CALIFORNIA, SAN DIEGO

**UCSD**

---

## *Getting an allocation*

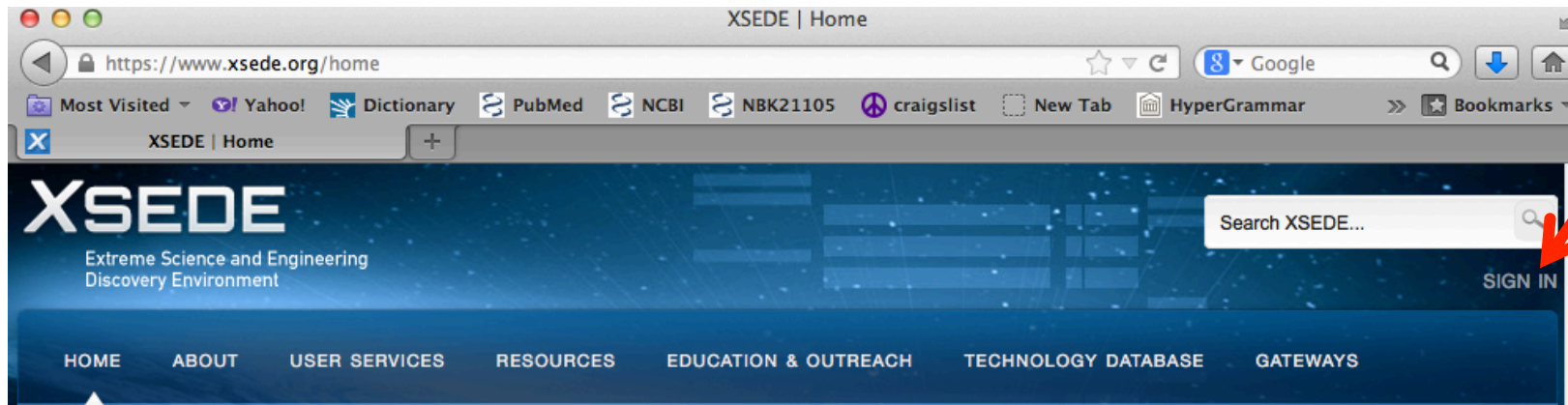
### *The WRONG way*

- Send email to someone you know at SDSC
- Who will then forward your email to me asking if I can help a new user who wants to get started with SDSC or XSEDE resources
- I'll then send you an email explaining what XSEDE is and why you need to go through the standard allocations channels

### *The RIGHT way*

- Register at the XSEDE portal and submit your proposal through POPS

## Allocations step 1 – create a portal account



USER PORTAL

USER NAME

PASSWORD

**SIGN IN**  REMEMBER ME

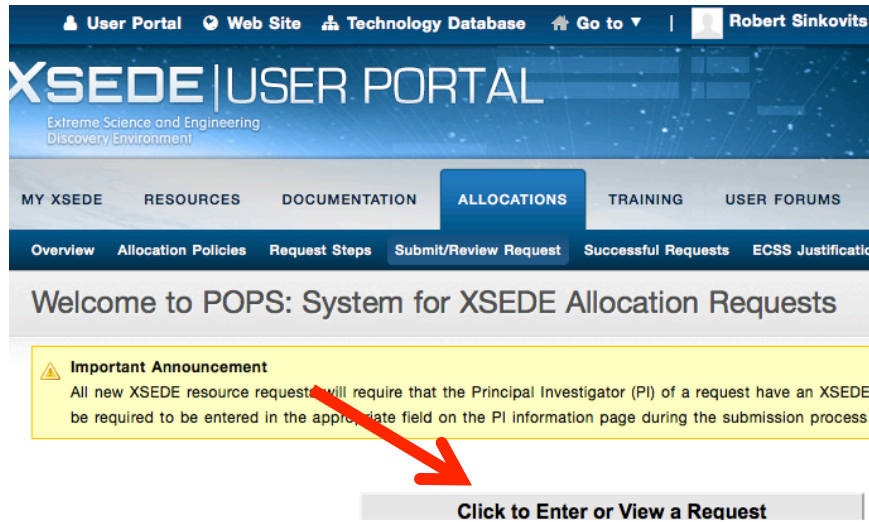
[Create Account](#) | [Verify Account](#) | [Forgot Password](#)

All users must create a portal account in order to use any of the XSEDE resources



## Allocations step 2 – Navigate to POPS

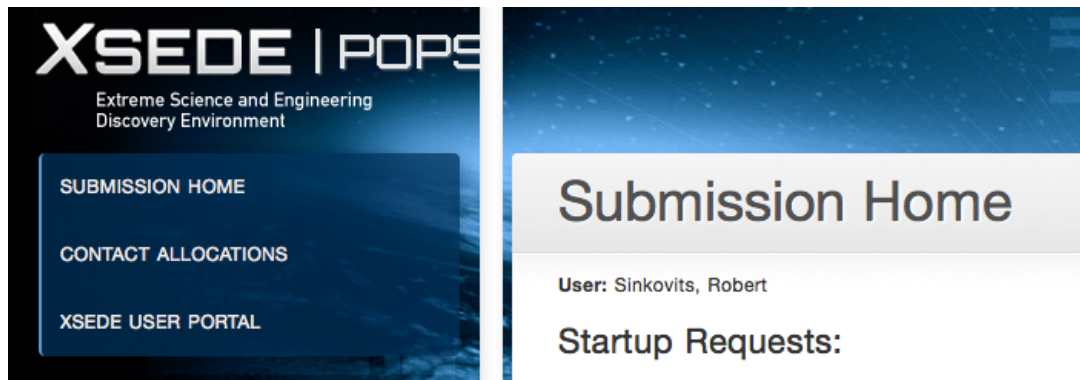
<https://www.xsede.org/group/xup/submit-request>



The screenshot shows the XSEDE User Portal interface. At the top, there are navigation links for User Portal, Web Site, and Technology Database, along with a user profile for Robert Sinkovits. The main header reads "XSEDE | USER PORTAL" with the tagline "Extreme Science and Engineering Discovery Environment". Below this is a menu with categories: MY XSEDE, RESOURCES, DOCUMENTATION, ALLOCATIONS (highlighted), TRAINING, and USER FORUMS. Under the ALLOCATIONS category, there are sub-links: Overview, Allocation Policies, Request Steps, Submit/Review Request, Successful Requests, and ECSS Justification. A welcome message says "Welcome to POPS: System for XSEDE Allocation Requests". A yellow box contains an "Important Announcement" stating that all new XSEDE resource requests will require the Principal Investigator (PI) to have an XSEDE ID, which must be entered in the appropriate field on the PI information page. A red arrow points from this announcement to a button labeled "Click to Enter or View a Request".

Only the PI needs to apply for an allocation. The PI can then add an arbitrary number of users to the project.

Students cannot serve as PIs, except for NSF Graduate Research Fellows and Honorable Mention awardees.



The screenshot shows the "Submission Home" page for XSEDE POPS. The header includes "XSEDE | POPS" and "Extreme Science and Engineering Discovery Environment". On the left, there is a sidebar with links for "SUBMISSION HOME", "CONTACT ALLOCATIONS", and "XSEDE USER PORTAL". The main content area displays "Submission Home" and the user information "User: Sinkovits, Robert". Below this, it says "Startup Requests:".

# Getting help through XSEDE

## The RIGHT way

<https://www.xsede.org/web/xup/help-desk>

*All fields are required unless otherwise indicated*

FIRST NAME

Robert

LAST NAME

Sinkovits

EMAIL

sinkovit@sdsc.edu

CC (OPTIONAL)

*(Sends a carbon-copy of this update to a comma-delimited list of email addresses.)*

CATEGORY

Accounts

SYSTEM (OPTIONAL)

Blacklight  
Data Supercell  
futuregrid0.futuregrid.xsede  
Gordon Compute Cluster  
Gordon ION  
HPSS  
Keeneland  
Keeneland-KIDS

SUBJECT

PROBLEM DESCRIPTION

## The WRONG way

Send email to someone you know at SDSC, who will

Forward to someone who they think can help, who will eventually tell you to submit a ticket through the XSEDE ... or write your problem on a sticky note, which gets forgotten or falls behind desk

Requests submitted to the XSEDE ticketing system will be logged, tracked and assigned to the person best suited to handle your problem

---

## Moving time between XSEDE systems

### The RIGHT way

[https://pops-submit.xsede.org/auth/TGUP\\_POPS/main/cgi/index.cgi](https://pops-submit.xsede.org/auth/TGUP_POPS/main/cgi/index.cgi)

## Submission Home

User: Sinkovits, Robert

### Startup Requests:

1. New

IRI130008 Submission Status: **Approved A**

Title: **Predictive analytics gateway**

Submitted On: **Mar 22, 2013** Award Start

Actions: [[View](#) | [Transfer](#) | [Supplemental](#)]



### The WRONG way

Send email to someone you know at SDSC, who will ...

---

## ***Extended Collaborative Support Services (ECSS)***

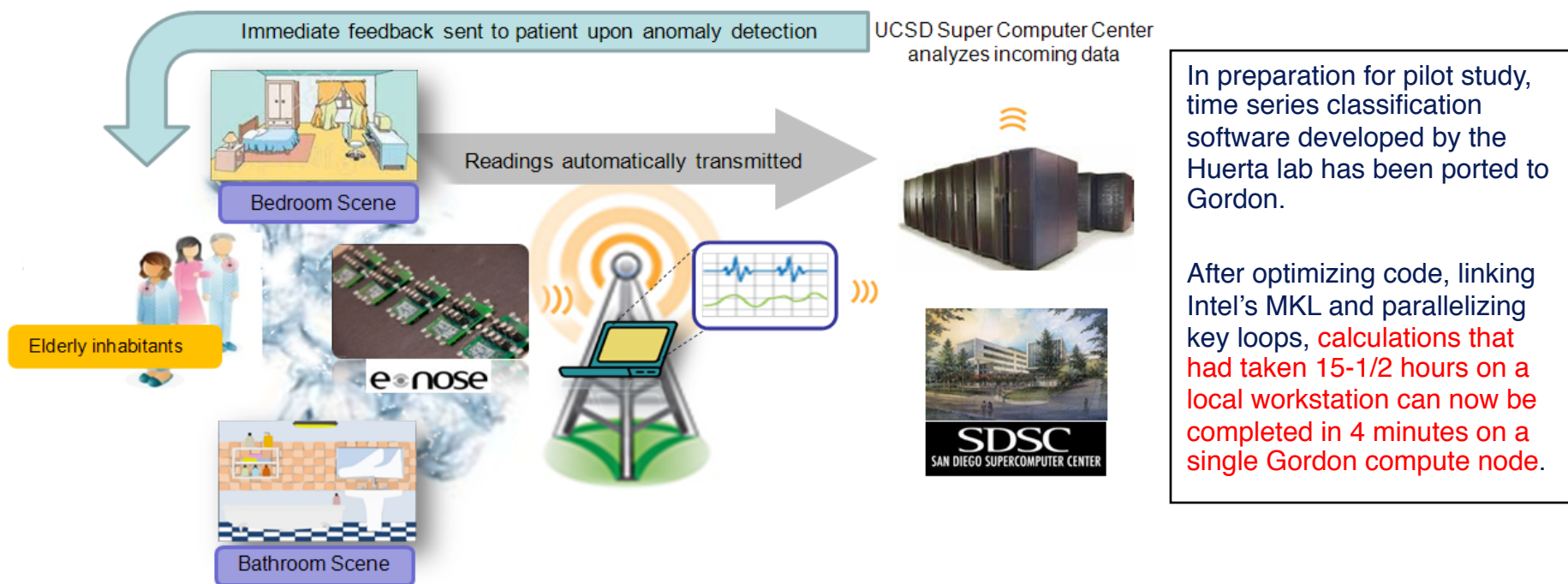
XSEDE provides much more than storage and compute cycles. Through the ECSS program, you can get help from expert staff to

- Make the transition from workstation to supercomputer
- Develop parallel versions of serial codes
- Optimize performance to make best use of XSEDE hardware
- Develop science gateways that allow entire communities of users to transparently access supercomputing resources
- Create workflows or other solutions that maximize throughput

Users normally ask for ECSS when they submit their allocations proposals, but it can be requested at any time

## ECSS example - classification of time series data

Chemical sensors (e-noses) will be placed in the homes of elderly participants in an effort to continuously and non-intrusively monitor their living environments. Time series classification algorithms will then be applied to the sensor data to detect anomalous behavior that may suggest a change in health status.



Source: Herb Hauser (U. Scranton) and Ramon Huerta (UCSD) Used by permission. 2012

**SDSC**

2013 Summer Institute: Discover Big Data, August 5-9, San Diego, California

SAN DIEGO SUPERCOMPUTER CENTER at the UNIVERSITY OF CALIFORNIA, SAN DIEGO

**UCSD**

## ***ECSS example - classification of time series data***

Original version of code was serial, compiled using GNU C++ compiler and linked to default LAPACK libraries. By changing the compiler and compiler options, linking MKL, enabling threaded execution, eliminating redundant calculations and parallelizing loops, obtained 167x speedup

Notes	cores	Run time	Speedup
Original code, GNU compiler	1	11:22:00	-
Switch to Intel compiler and enable AVX	1	05:41:49	2.0
Link threaded MKL library, run in parallel	16	00:14:46	46.2
OpenMP directives in loops in kAR and kARtest	16	00:13:10	52.5
Remove duplicate call to kARtest	16	00:07:58	85.6
Optimization of DYSRK operations	16	00:04:04	167.7

Reported speedups are relative to single core on Gordon. Porting from Huerta lab workstation (Intel Nehalem) to Gordon resulted in 1.3x reduction in runtime

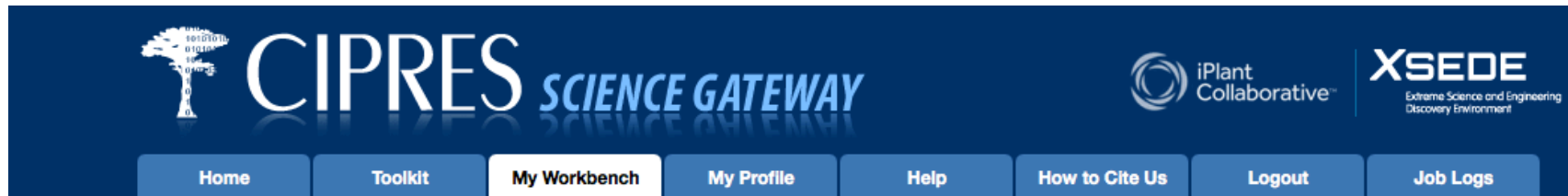
## Gateways

Gateways allow users to submit jobs that will subsequently be run on XSEDE resources through a web interface. In most cases a single allocation to the PI serves an entire community of users.

<https://www.xsede.org/web/guest/gateways-listing>


TITLE	FIELD OF SCIENCE	PORTAL HOMEPAGE
<a href="#">Massive Pulsar Surveys using the Arecibo L-band Feed Array (ALFA)</a>	Astronomical Sciences	<a href="#">Visit Portal</a>
<a href="#">Center for Multiscale Modeling of Atmospheric Processes</a>	Atmospheric Sciences	<a href="#">Visit Portal</a>
<a href="#">Community Climate System Model (CCSM) TeraGrid Gateway</a>	Atmospheric Sciences	<a href="#">Visit Portal</a>
<a href="#">Biodrugscore: A portal for customized scoring and ranking of molecules docked to the human proteome</a>	Biochemistry and Molecular Structure and Function	<a href="#">Visit Portal</a>
<a href="#">Chemical Informatics and Cyberinfrastructure Collaboratory</a>	Biochemistry and Molecular Structure and Function	<a href="#">Visit Portal</a>

# Gateway example - CIPRES



The header features the CIPRES Science Gateway logo on the left, the iPlant Collaborative logo in the center, and the XSEDE logo on the right. Below the logos is a navigation menu with buttons for Home, Toolkit, My Workbench (selected), My Profile, Help, How to Cite Us, Logout, and Job Logs.

## Folders



Folder icons for a file system, with a label 'GPCR' and sub-items 'Data (0)' and 'Tasks (0)'.



Workflow tabs: Task Summary, Select Data, Select Tool, and Set Parameters.

BEAST on XSEDE: Bayesian Evolutionary Analysis by Sampling Trees - run on XSEDE ([John P. Huelsenbeck and Fred Ronquist](#))

[Simple Parameters](#) OPEN / CLOSE

Maximum Hours to Run (up to 334 hours) \*

My data set is partitioned \*

How many partitions does your data have? \*

Do not use Beagle

Specify a seed for this run (by default a random seed is used)

Enter the seed value here

Set the Beagle Scaling Parameter (the choice can affect performance) \*

*CIPRES gateway enables users to specify data sets, software and parameters through web interface. Job is then launched on SDSC's Gordon or Trestles systems*



---

## *XSEDE Wrap up*

- XSEDE centralizes many of the essential service. Users do not need to figure out a different set of policies, web sites, etc. in order to work across multiple sites.
- Get used to going through the proper XSEDE channels for allocations, transfers, adding users to accounts and submitting help tickets
- XSEDE is much more than hardware. Our staff have expertise in a wide range of higher performance computing topics.