OSG

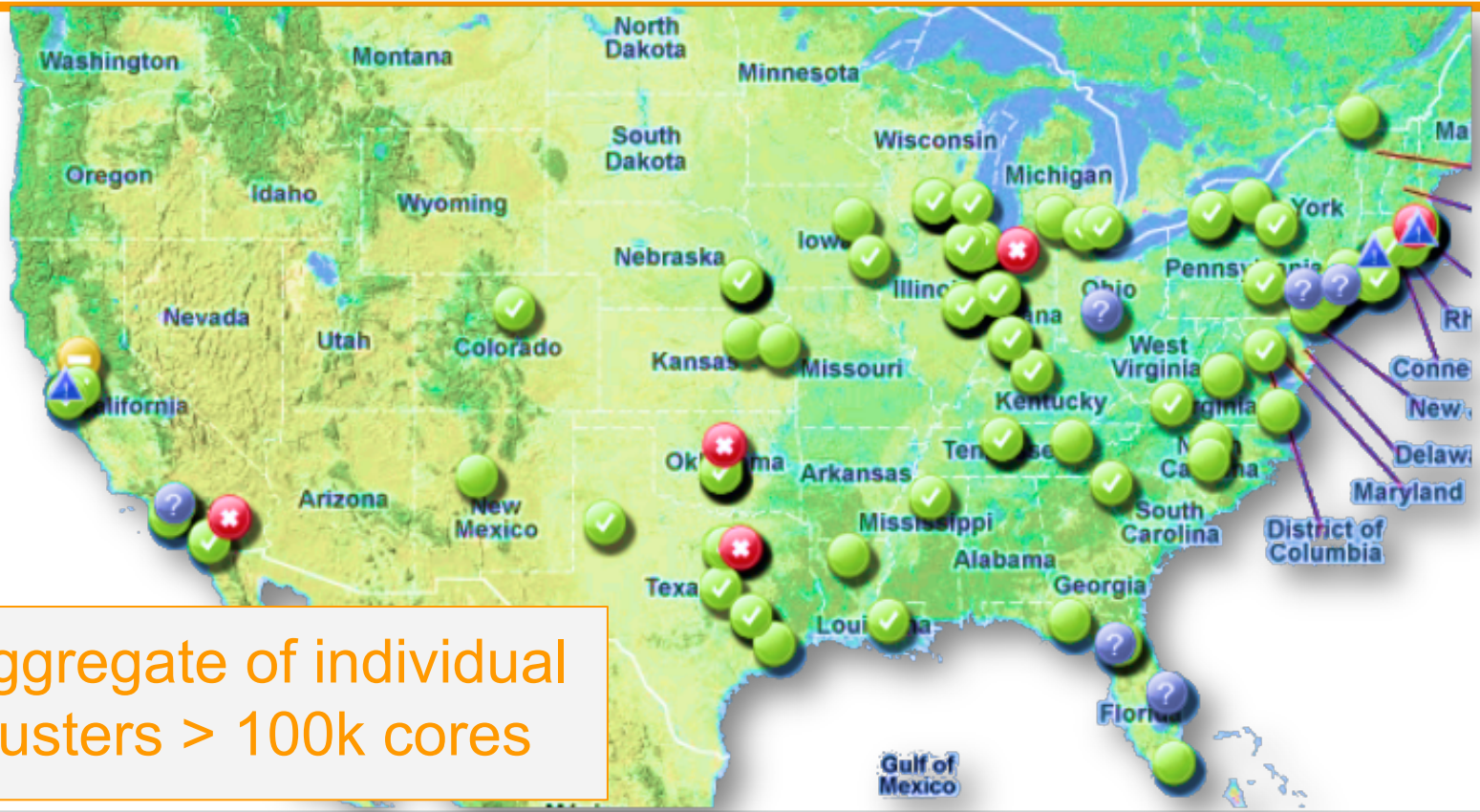
Advanced Cyberinfrastructure Research &
Education Facilitators Virtual Residency
Oklahoma University, June 4, 2015

Open Science Grid

Rob Gardner • University of Chicago

OSG User Support and Campus Grids area coordinator
rwg@uchicago.edu

Open Science Grid: **HTC supercomputer**



aggregate of individual clusters > 100k cores

Open Science Grid: **HTC supercomputer**

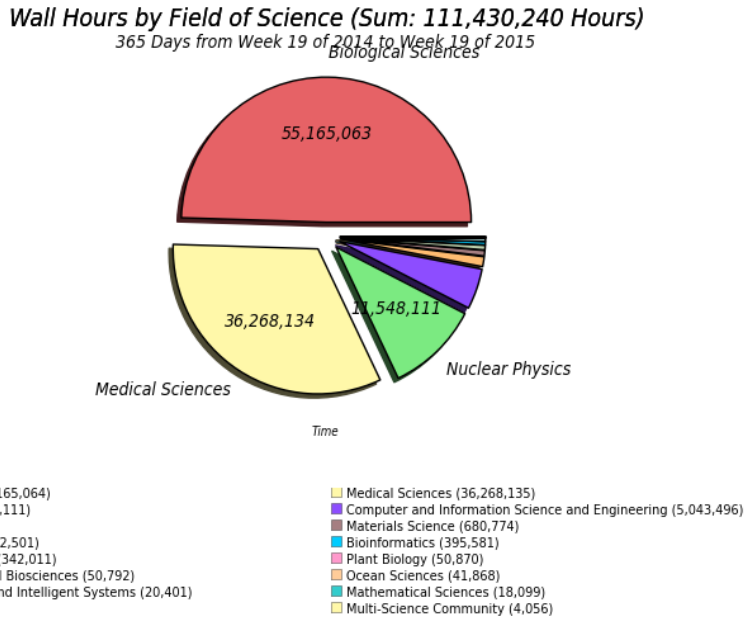
- 2014 stats
  - **67%** size of XD, **35%** BlueWaters
  - **2.5 Million CPU hours/day**
  - **800M hours/year**
  - **125M/y** provided **opportunistic**
- **>1 petabyte** of data/day
- **50+** research groups
- **thousands** of users
- XD service provider for XSEDE

Network	Billion CPU-Hours provided in 2014
XDNet	1.2
BlueWaters	2.3
NCAR	0.57
OSG	0.8

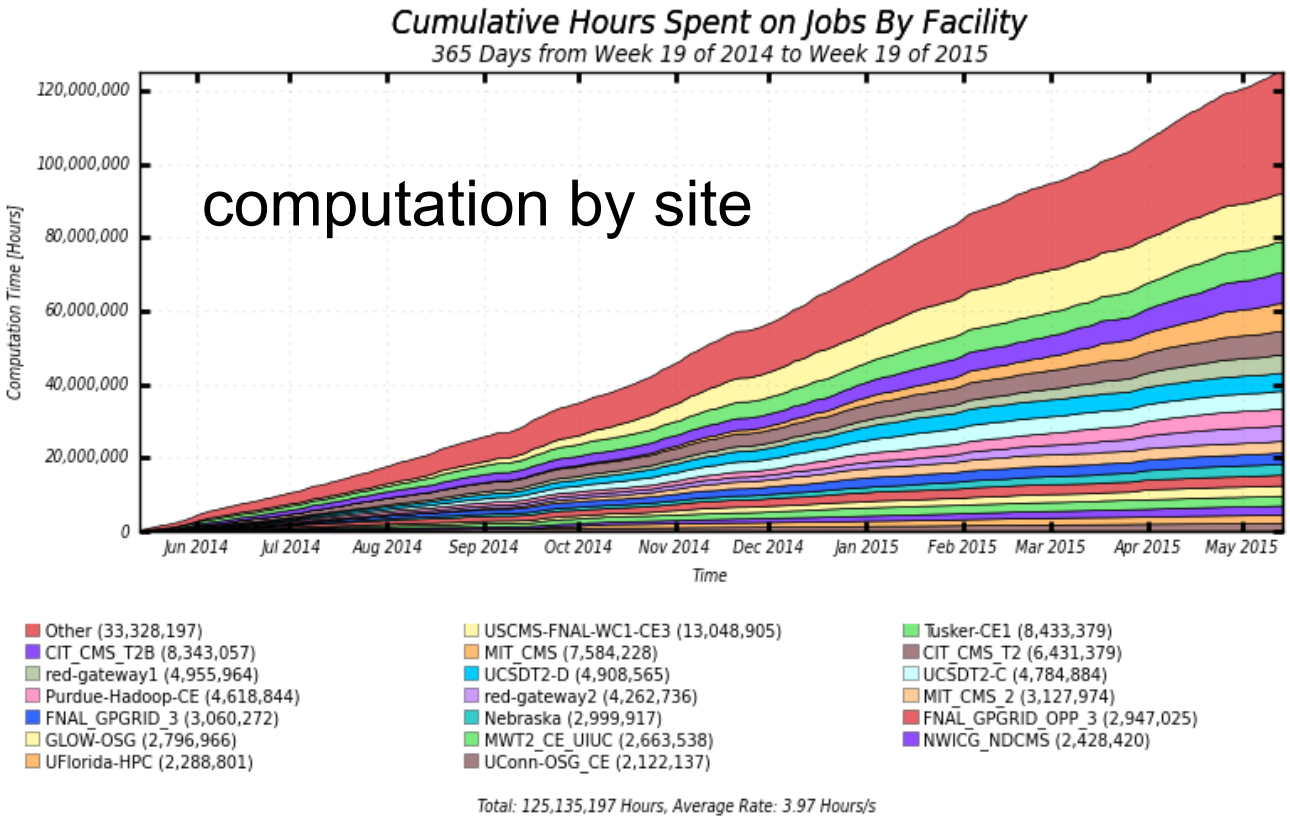
Rudi Eigenmann
Program Director Division of
Advanced Cyberinfrastructure (ACI)
NSF CISE
CASC Meeting, April 1, 2015

Science that has benefited

- Biological & medical sciences
- Nuclear theory
- Chemistry
- Molecular dynamics
- Materials
- Genetics
- Drug studies



OSG virtual organization

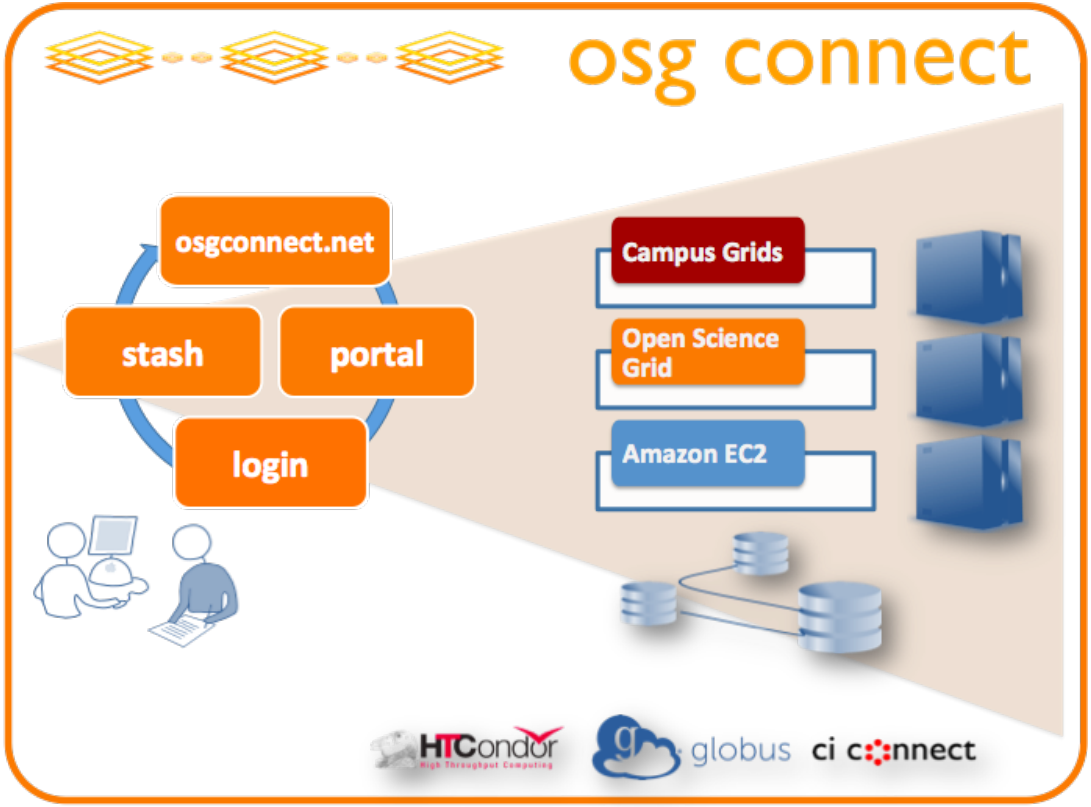


125 million hours opportunistic for individual projects

Is HTC for you?

- Does your science workflow involve computations that can be split into many independent jobs?
- Can individual jobs be run on a single processor (as opposed to a single large-scale MPI job simultaneously utilizing many processors)?
- Can your applications be interrupted and restarted at a later time or on another processor or computing site?
- Is your application software “portable”?
- How does computing fit into your science workflow? Are results needed immediately after submission?
- Do your computations require access to or produce huge amounts of data?

OSG Connect Service



OSG as a **campus cluster**

- ★ Login host
- ★ Job scheduler
- ★ Software
- ★ Storage

Submit jobs to OSG with HTCondor

- Simple HTCondor submission
- Complexity hidden from the user
- No grid certificates required
- Uses HTCondor ClassAd and glidein technology
- DAGMan and other workflow tools

Software & tools on the OSG

- Distributed software file system
- Special **`module`** command
  - *identical software on all clusters*

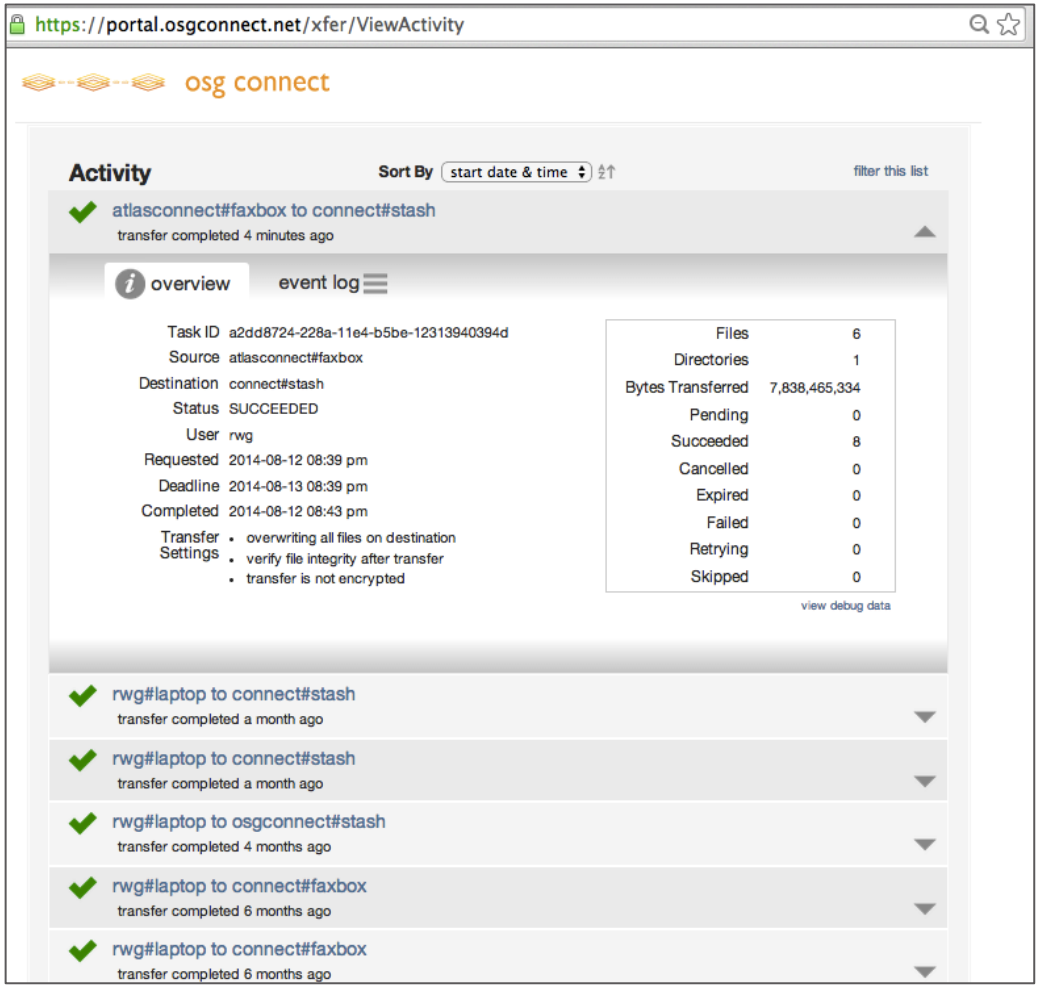
```
[rwg@login01 ~]$ module avail

-------------------- /cvmfs/oasis.opensciencegrid.org/osg/modules/modulefiles/Core -------------
   R/3.1.1              ectools              hdf5/1.8.13      (D)    pbsuite/14.9.9
   SitePackage          espresso/5.1         hmmer/3.1               pcre/8.35
   SparseSuite/4.2.1    fftw/3.3.4-gromacs   jpeg                    protobuf/2.5
   ant/1.9.4            fftw/3.3.4       (D) lammps/2.0              python/2.7       (D)
   apr/1.5.1            fpc/2.6.4            lapack                  python/3.4
   apr-util/1.5.3       gamess/2013          lmod/5.6.2              qhull/2012.1
   aprutil/1.5.3        gcc/4.6.2            madgraph/2.1.2          root/5.34-21
   atlas                geos/3.4.2           mafft/7.1               samtools/0.1.17
   autodock/4.2.6       git/1.9.0            matlab/2013b            serf/1.37
   bedtools/2.21        glpk/4.54            matlab/2014a     (D)    settarg/5.6.2
   blasr/1.3.1          gnuplot/4.6.5        mercurial/1.9.1         siesta/3.2
   blast                graphviz/2.38.0      mplayer/1.1             subversion/1.8.10
   blender              gromacs/4.6.5        mrbayes/3.2.2           uclust/2.22
   cmake/3.0.1          gromacs/5.0.0    (D) muscle/3.8.31           valgrind/3.10
   cp2k/2.5.1           gsl/1.16             namd/2.9                vmd/1.9.1
   curl/7.37.1          hdf5/1.8.12          octave/3.8.1            wget/1.15
```

```
$ switchmodules oasis
$ module avail
$ module load R
$ module load namd
$ module purge
$ switchmodules local
```

- Common tools & libs
- Curate on demand continuously
- HTC apps in **XSEDE campus bridging yum repo**

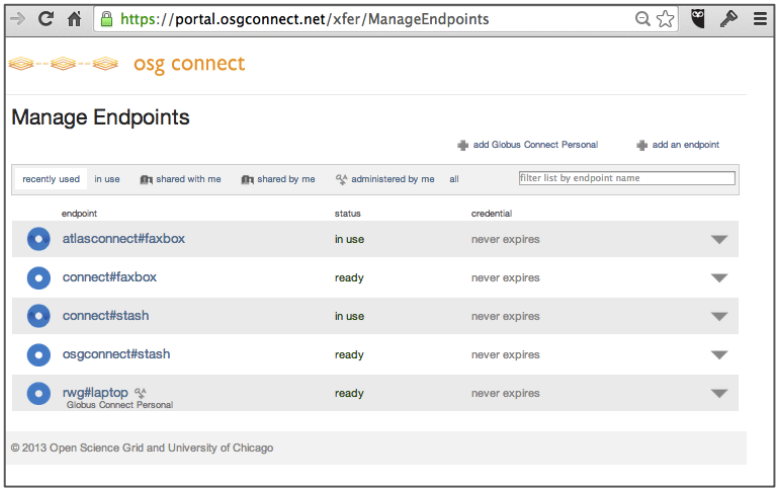
Storage service: “Stash”

- Provide a quasi-transient storage service for job input/output data
- **POSIX** access provided to the login host
- **Globus** Online Server for managed transfers from campus data services
- Personalized **http** service endpoint
- Can handle data sets up to **10 TB/user**
- Connected to 100 Gbps SciDMZ (I2, ESnet)



Easy, reliable upload of data to Stash for processing on OSG, and downloading output data back to campus.

globus



Quotable: Searching for Symmetry Orbits & Rainbow Cliques in Group Theory

From: Anton Betten <betten@math.colostate.edu>

Date: Wed, Sep 3, 2014 at 10:47 AM

Subject: OSG runs

Hi Suchandra,

just a quick note that **everything is going extremely well with my computations on the OSG. I start 10K jobs one day and the next day they are done. Fantastic!**

I am probably more than 50% done by now, and I am trying hard to complete before my next conference in Germany (in less than two weeks). If I can report that the classification is finished and confirm the published result, that would be great.

Thanks, Anton

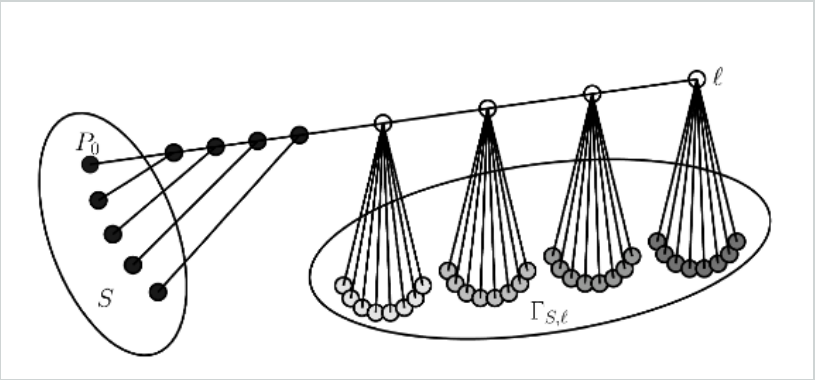
Anton Betten, Ph.D.

Associate Professor

Department of Mathematics

Colorado State University

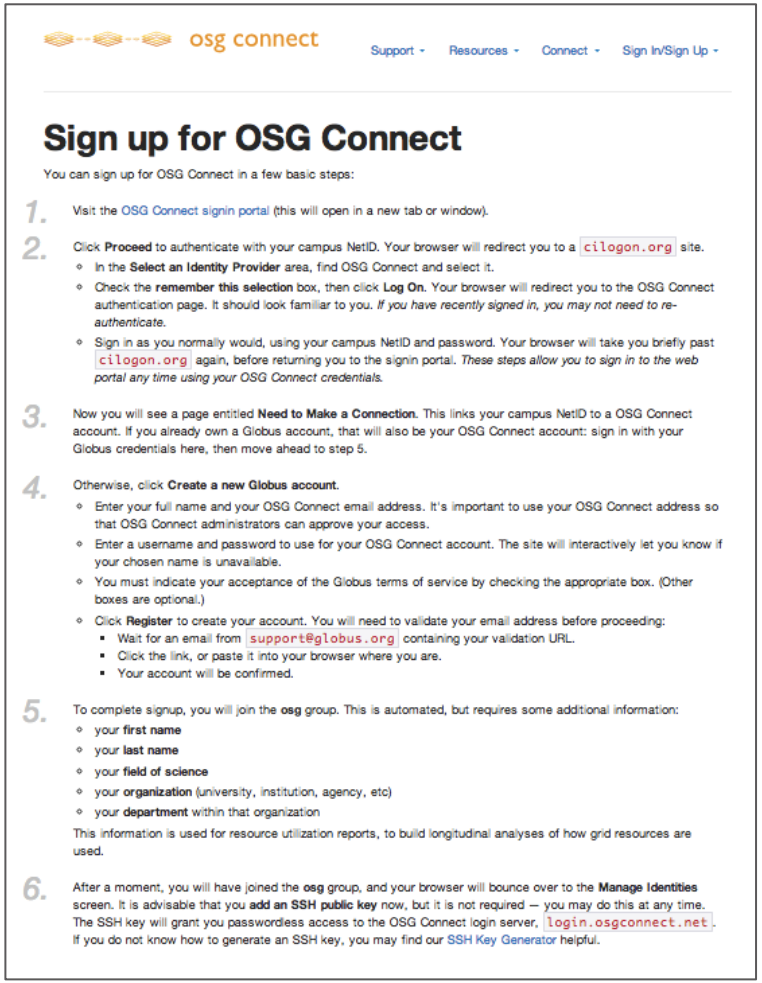
Fort Collins, CO 80523-1874, U.S.A.



“My recommendation for other scientists looking into using the OSG is don’t be intimidated, find someone to work with, and don’t hesitate to reach out… The competence and concern of the staff, and their ability to communicate with non-computer scientists are a huge help. It seems like OSG really has a service mentality.”

Patrick Reeves, The National Center for Genetic Resources Preservation (NCGRP), U.S. Department of Agriculture

http://osgconnect.net/



osg connect

Support Resources Connect Sign In/Sign Up

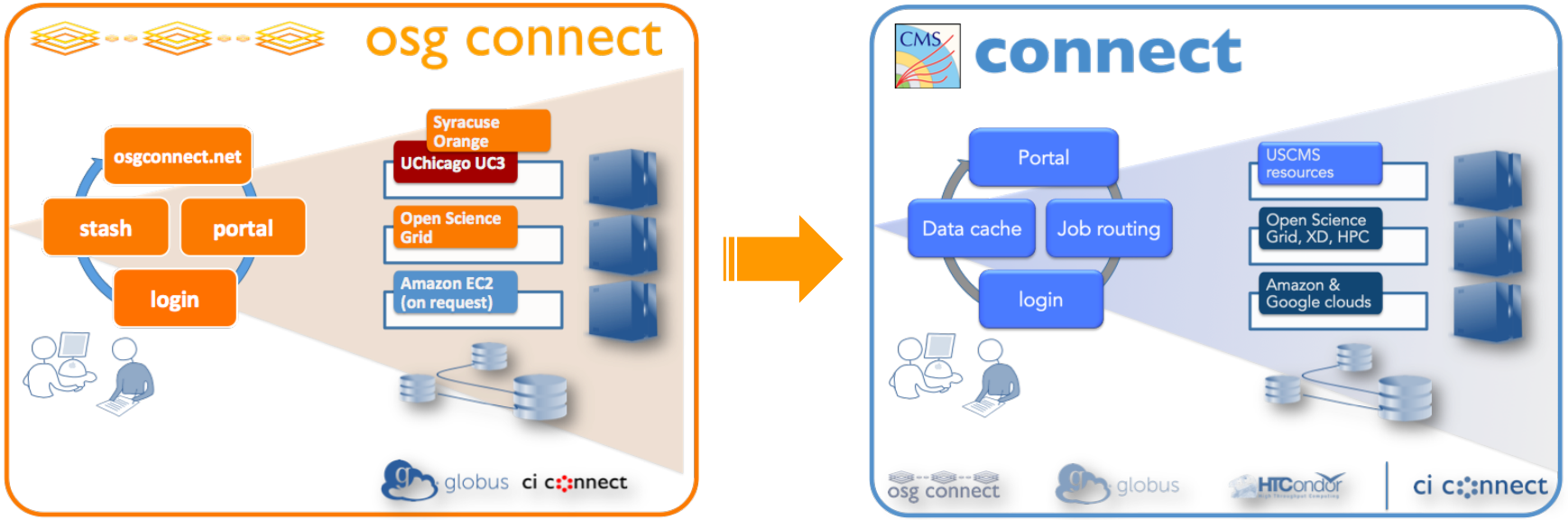
Sign up for OSG Connect

You can sign up for OSG Connect in a few basic steps:

1. Visit the OSG Connect signin portal (this will open in a new tab or window).
2. Click **Proceed** to authenticate with your campus NetID. Your browser will redirect you to a `cilogon.org` site.
   - In the **Select an Identity Provider** area, find OSG Connect and select it.
   - Check the **remember this selection** box, then click **Log On**. Your browser will redirect you to the OSG Connect authentication page. It should look familiar to you. *If you have recently signed in, you may not need to re-authenticate.*
   - Sign in as you normally would, using your campus NetID and password. Your browser will take you briefly past `cilogon.org` again, before returning you to the signin portal. *These steps allow you to sign in to the web portal any time using your OSG Connect credentials.*
3. Now you will see a page entitled **Need to Make a Connection**. This links your campus NetID to a OSG Connect account. If you already own a Globus account, that will also be your OSG Connect account: sign in with your Globus credentials here, then move ahead to step 5.
4. Otherwise, click **Create a new Globus account**.
   - Enter your full name and your OSG Connect email address. It's important to use your OSG Connect address so that OSG Connect administrators can approve your access.
   - Enter a username and password to use for your OSG Connect account. The site will interactively let you know if your chosen name is unavailable.
   - You must indicate your acceptance of the Globus terms of service by checking the appropriate box. (Other boxes are optional.)
   - Click **Register** to create your account. You will need to validate your email address before proceeding:
     - Wait for an email from `support@globus.org` containing your validation URL.
     - Click the link, or paste it into your browser where you are.
     - Your account will be confirmed.
5. To complete signup, you will join the **osg** group. This is automated, but requires some additional information:
   - your **first name**
   - your **last name**
   - your **field of science**
   - your **organization** (university, institution, agency, etc)
   - your **department** within that organization

   This information is used for resource utilization reports, to build longitudinal analyses of how grid resources are used.
6. After a moment, you will have joined the **osg** group, and your browser will bounce over to the **Manage Identities** screen. It is advisable that you **add an SSH public key** now, but it is not required — you may do this at any time. The SSH key will grant you passwordless access to the OSG Connect login server, `login.osgconnect.net`. If you do not know how to generate an SSH key, you may find our SSH Key Generator helpful.

Communities or campuses, as a service



CI Connect - services for campus grids (Duke, UChicago)
CMS Connect (users from 50 schools) - led by Notre Dame

Summary and Conclusions

- For **researchers:**
  - OSG Connect provides a quick "**on-ramp**" to resources in the Open Science Grid
- For **campus providers:**
  - OSG Connect offers a service **to connect campus HPC centers** to the OSG in lightweight fashion
  - Hosted "Compute Element" available for larger sites
- For **communities & campus grids**: CI Connect to **share** resources and collaborate easily

opensciencegrid

@rwg