Open Science Grid

Introduction to the Open Science Grid and the OSG Match Maker

6/20/10 13:15

Mats Rynge <rynge@isi.edu>

OSG Engagement Team

USC Information Sciences Institute

The Open Science Grid

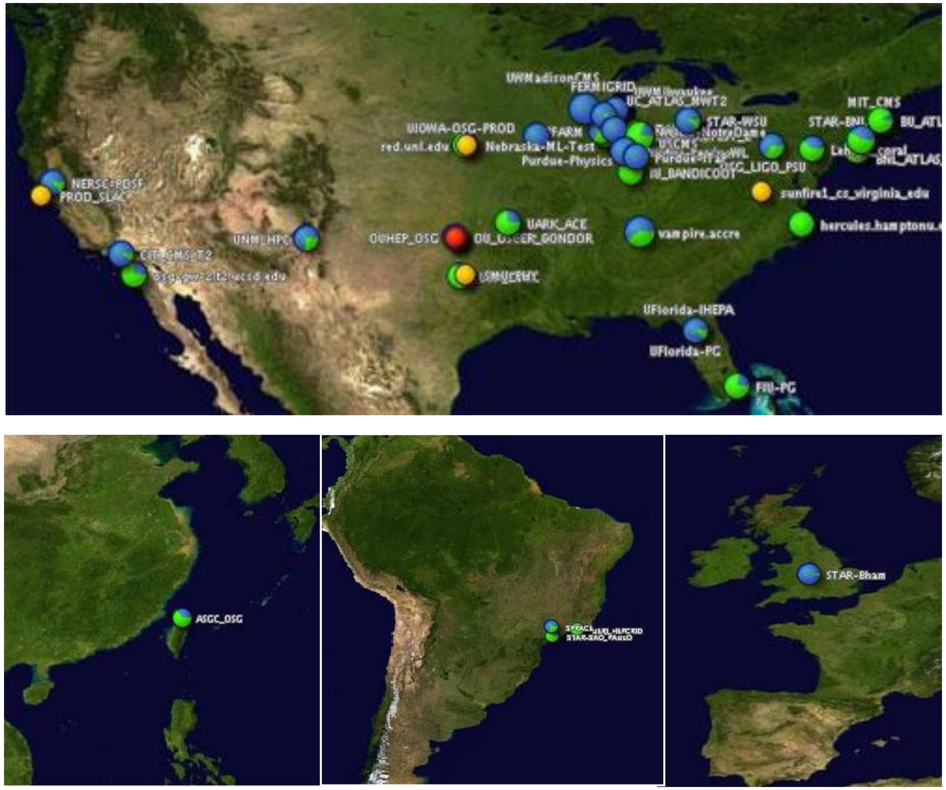
A framework for large scale distributed resource sharing

addressing the technology, policy, and social requirements of sharing

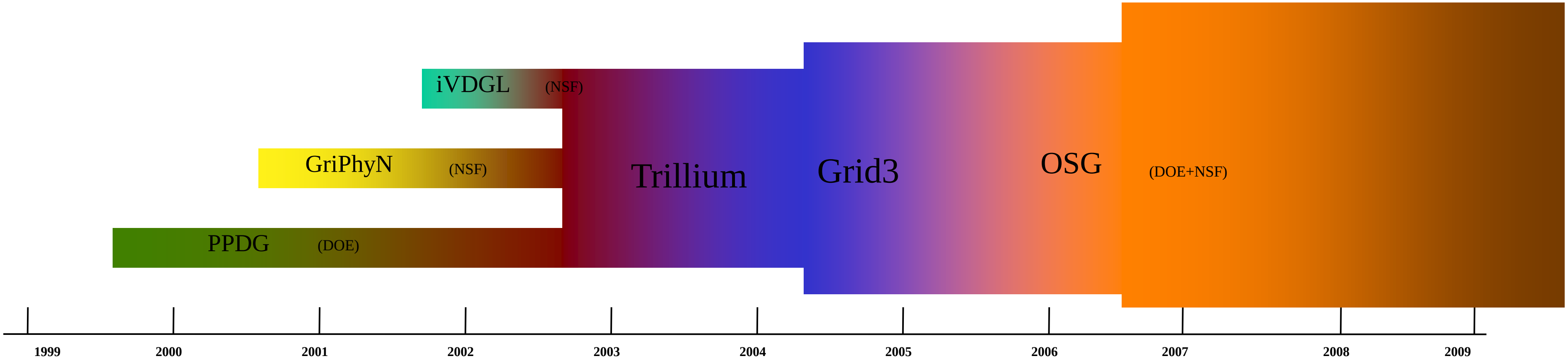
OSG is a consortium of software, service and resource providers and researchers, from universities, national laboratories and computing centers across the U.S., who together build and operate the OSG project. The project is funded by the NSF and DOE, and provides staff for managing various aspects of the OSG.

Brings petascale computing and storage resources into a uniform grid computing environment

Integrates computing and storage resources from over 80 sites in the U.S. and beyond

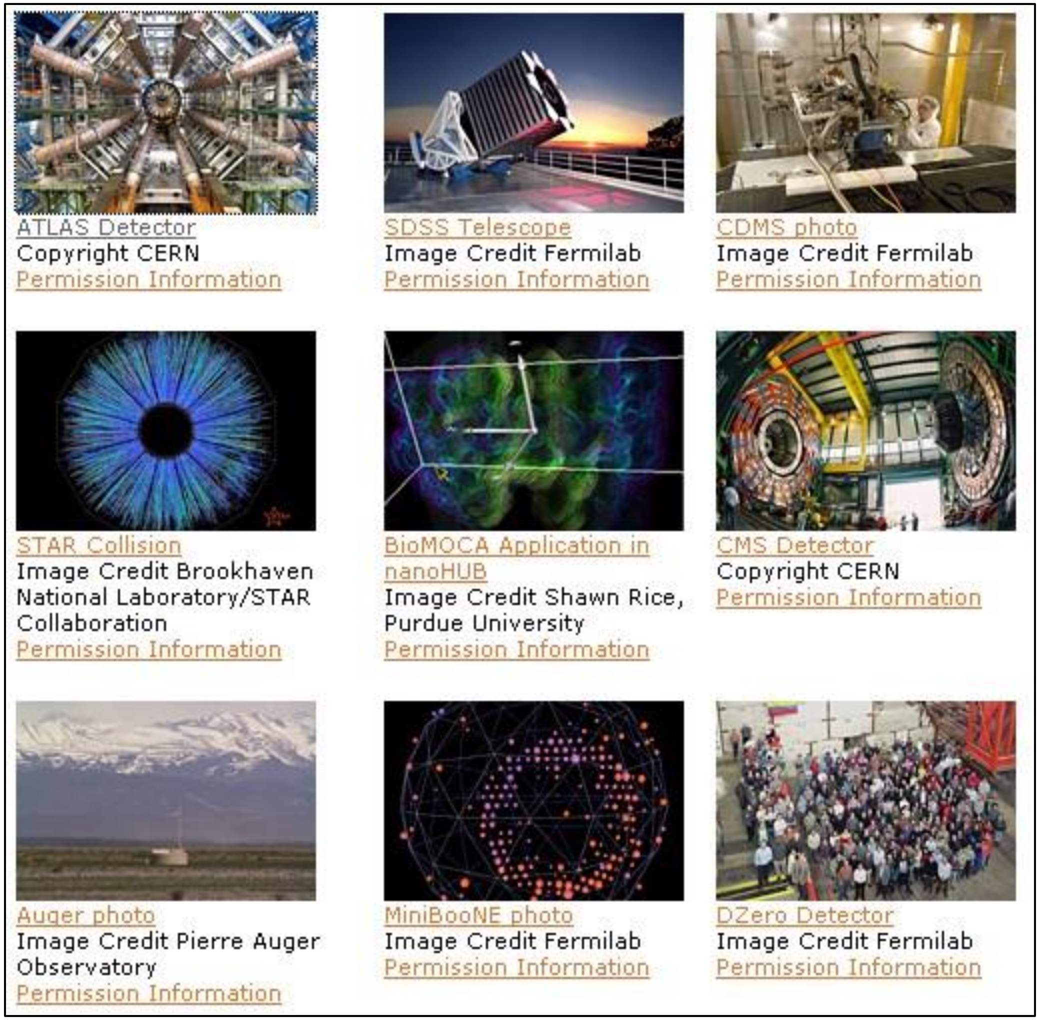


Context: Evolution of Projects



Using OSG Today

- Astrophysics
- Biochemistry
- Bioinformatics
- Earthquake Engineering
- Genetics
- Gravitational-wave physics
- Mathematics
- Nanotechnology
- Nuclear and particle physics
- Text mining
- And more…



ATLAS Detector
Copyright CERN
Permission Information

SDSS Telescope
Image Credit Fermilab
Permission Information

CDMS photo
Image Credit Fermilab
Permission Information

STAR Collision
Image Credit Brookhaven National Laboratory/STAR Collaboration
Permission Information

BioMOCA Application in nanoHUB
Image Credit Shawn Rice, Purdue University
Permission Information

CMS Detector
Copyright CERN
Permission Information

Auger photo
Image Credit Pierre Auger Observatory
Permission Information

MiniBooNE photo
Image Credit Fermilab
Permission Information

DZero Detector
Image Credit Fermilab
Permission Information

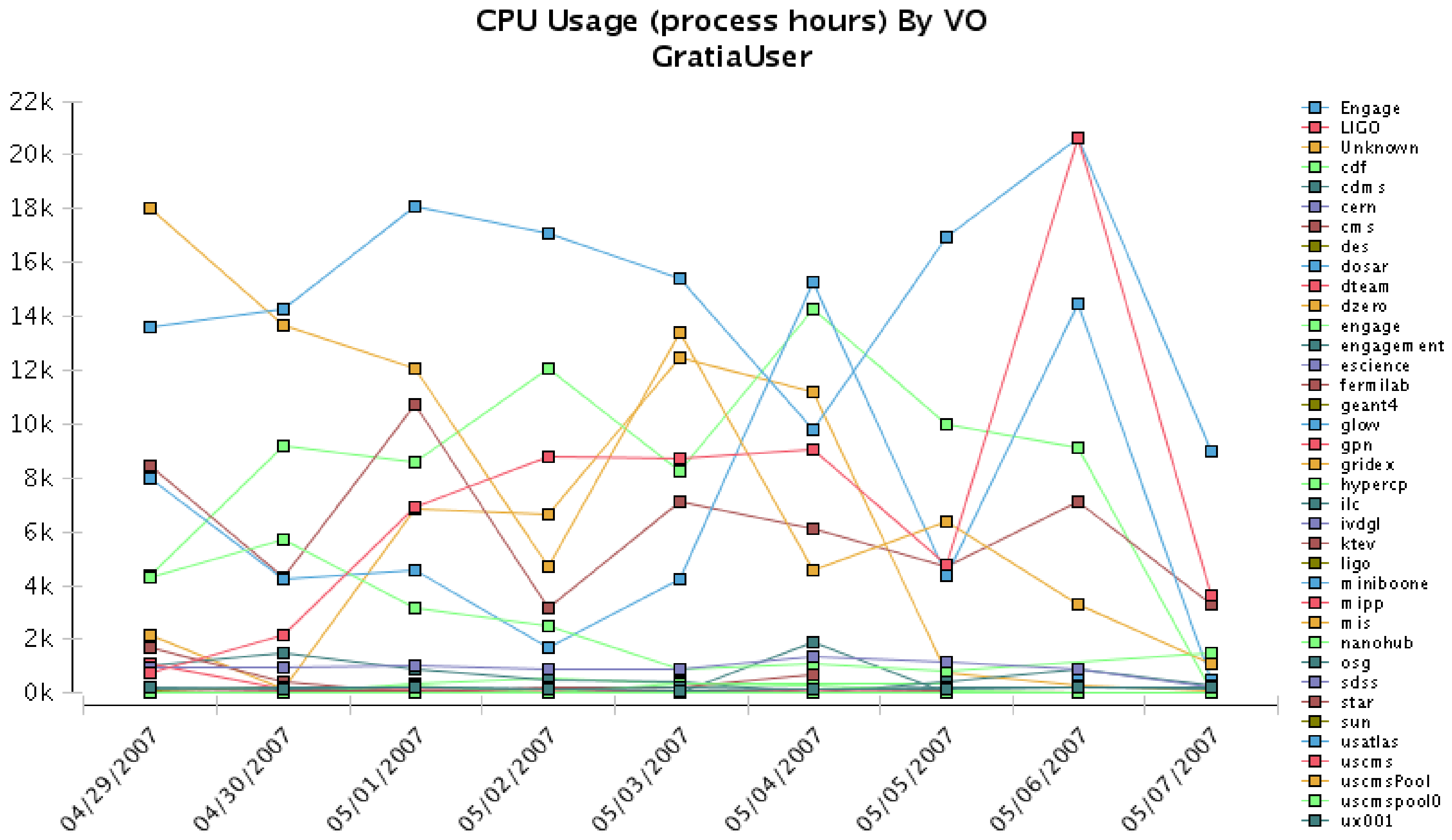
OSG Engagement Mission

- Help new user communities from diverse scientific domains adapt their computational systems to leverage OSG

- Facilitate University Campus CI deployment, and interconnect it with the national community

- Provide feedback and new requirements to the infrastructure providers

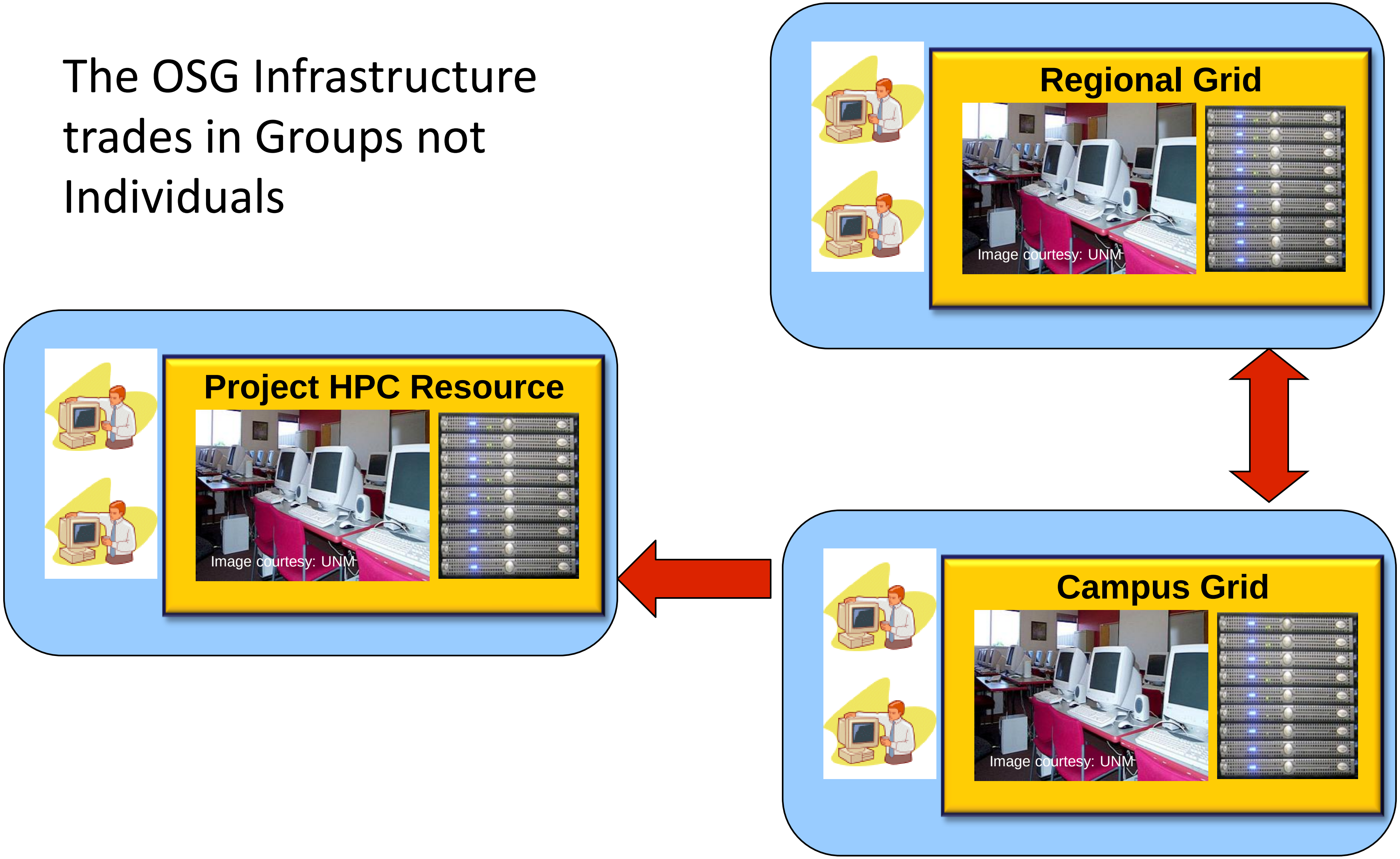
Opportunistic Cycles



Date range: 2007-04-29 00:00:00 GMT - 2007-05-07 23:59:59 GMT

Virtual Organizations (VOs)

The OSG Infrastructure trades in Groups not Individuals



Workload Management Systems (WMS)

- Condor-G
- OSG Match Maker
  - Condor-G + site selection
- glidinWMS
  - Condor Glideins
- PanDA
  - Custom pilots

OSGMM – OSG Match Maker

- Simple match maker for Condor-G jobs
  - Based on *“Matchmaking in the Grid Universe”* in the Condor manual
- Open Source
  - http://osgmm.sourceforge.net/
- Installs on top of the OSG Client software stack

What is Resource Selection?

- Well described jobs and resources

- Can you list all the requirements for your jobs?
  - Memory usage? Disk usage? Dependencies?

- **Automatically** match the jobs up against resources

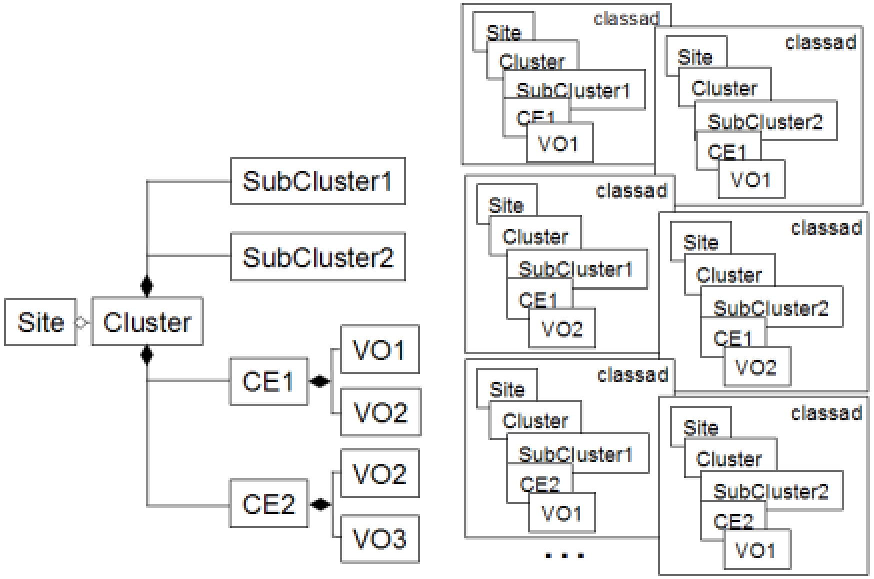
- Additional features include
  - automatic retries of failed jobs
  - site verification

OSG: Resource Discovery

- CE advertises capabilities and state (GIP & CEMon)
- ReSS - Resource Selection Service
  - Condor ClassAd format
- BDII - Berkeley Database Information Index
  - LDIF format

ReSS

- Collects data from compute elements (CE), storage elements (SE), and software entities
- Publishes the data in Condor ClassAd format
- One ClassAd per Cluster, Subcluster, CE, SE, VO
  - Cardinality of CE*Cluster*Subcluster*VO*SE*VO
  - Currently about 15,000 ads



Information in ReSS

- OS name / version
- LRM information
  - Total number of job slots
  - Assigned slots
  - Open job slots
- Memory / CPU / Disk
- Network setup
- Storage configuration

•Validity of ClassAds

- Each ad augmented with validity tests in the form of classad attributes
- Test attributes are put in logical 'AND' in the attribute 'isClassadValid'

ReSS ClassAd

```
MyType = "Machine"
GlueSubClusterLogicalCPUs = 2
GlueCEPolicyAssignedJobSlots = 0
GlueCEInfoHostName = "antaeus.hpcc.ttu.edu"
GlueHostNetworkAdapterOutboundIP = TRUE
GlueHostArchitectureSMPSize = 2
OSGMM_Software_Rosetta_v3 = TRUE
OSGMM_MemPerCPU = 1010460
GlueSubClusterWNTmpDir = "/state/partition1"
OSGMM_OSGAPPWriteWorkNode = TRUE
GlueCEInfoContactString = "antaeus.hpcc.ttu.edu:2119/jobmanager-lsf"
GlueHostOperatingSystemName = "CentOS"
```

Retrieving Information from ReSS

```
COLLECTOR_HOST  = osg-ress-1.fnal.gov
HOSTALLOW_NEGOTIATOR = osg-ress-1.fnal.gov
HOSTALLOW_NEGOTIATOR_SCHEDD = original_value,
                                        osg-ress-1.fnal.gov
```

```
condor_status -any -constraint
    'StringlistIMember("VO:Engage";
    GlueCEAccessControlBaseRule)'
    -pool osg-ress-1.fnal.gov
```

Have OSGMM do it for you!

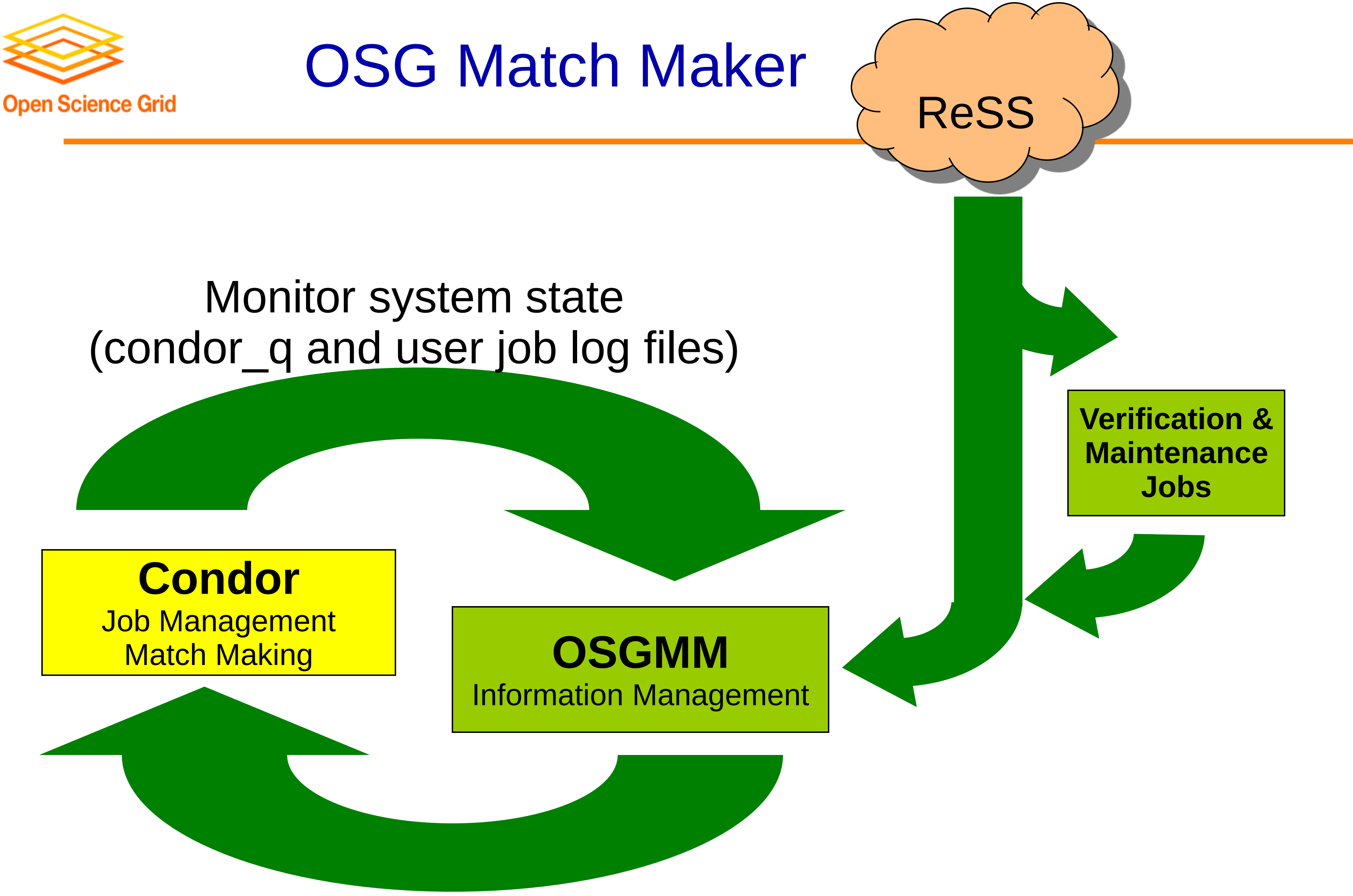
Open Science Grid

OSGMM – How does it work?

- Retrieve base ClassAds from ReSS
- Validate/maintain the sites with probe jobs
- Determine the current state of the system by looking at current job states and success rates (continuous system feedback)
- Merge the information, and insert into local Condor system
- Let Condor manage the jobs

Open Science Grid

OSG Match Maker



Site Rank

- Integer between 0 and 1000
- Calculated every minute from current state and some history
- Factors:
  - Jobs submitting/staging/pending/running provides the baseline
  - Job success rate for the site over the last 6 hours
  - Ratio between matched jobs, and the max number we want on that site

Periodic Hold/Release

- Job fails...
- Job is in the queue for too long...
- Job is running for too long...

resubmit to another site

- When submitting to another site, do not submit to a site which we have already failed on

Condor Submit File

```
globusscheduler = $$(GlueCEInfoContactString)

requirements = (
   (TARGET.GlueCEInfoContactString =!= UNDEFINED) &&
   (TARGET.Rank > 300) &&
   (TARGET.OSGMM_CENetworkOutbound == True) &&
   (TARGET.OSGMM_SoftwareGlobusUrlCopy == True) &
   (TARGET.OSGMM_MemPerCPU >= 500000) )
```

```
# when retrying, remember the last 4 resources tried
match_list_length = 4
Rank = (TARGET.Rank) -
   ((TARGET.Name =?= LastMatchName0) * 1000) -
   ((TARGET.Name =?= LastMatchName1) * 1000) -
   ((TARGET.Name =?= LastMatchName2) * 1000) -
   ((TARGET.Name =?= LastMatchName3) * 1000)
```

Condor Submit File (cont.)

```
# make sure the job is being retried and rematched
periodic_release = (NumGlobusSubmits < 10)
globusresubmit = (NumSystemHolds >= NumJobMatches)
rematch = True
globus_rematch = True
```

```
# only allow for the job to be queued or running for a while
# then try to move it
#  JobStatus==1 is pending
#  JobStatus==2 is running
periodic_hold = (
  ((JobStatus==1) && ((CurrentTime - EnteredCurrentStatus) >
   (5*60*60)) ||
  ((JobStatus==2) && ((CurrentTime - EnteredCurrentStatus) >
   (24*60*60)) )
```

Open Science Grid

CLI: condor_grid_overview

```
ID       Owner   Resource          Status    Time Sta Sub
======== ======= ================= ========= ========= ===
46381    rynge    (DAGMan)                    1:58:54
46382    rynge   GLOW              Running    1:55:43   1
46384    rynge   UWMilwaukee       Pending    1:57:04   1
46387    rynge   Nebraska          Running    1:00:43   1


Site          Jobs  Subm  Pend  Run   Stage  Fail   Rank
============= ===== ===== ===== ===== ===== =====
=====
ASGC_OSG        17     0     0    15     2     0    155
FNAL_GPFARM     14     4     0    10     0     0    720
GLOW            36     6     5    22     3     0    372
Nebraska        17     0     5    12     0     0    288
Purdue-Lear     15     4     0    10     1     0    372
TTU-ANTAEUS     15     2     0    11     2     0    372
Vanderbilt      45     4     4    37     0     0    350
```

Questions?

OSG Engagement VO

https://twiki.grid.iu.edu/twiki/bin/view/Engagement/WebHome

engage-team@opensciencegrid.org