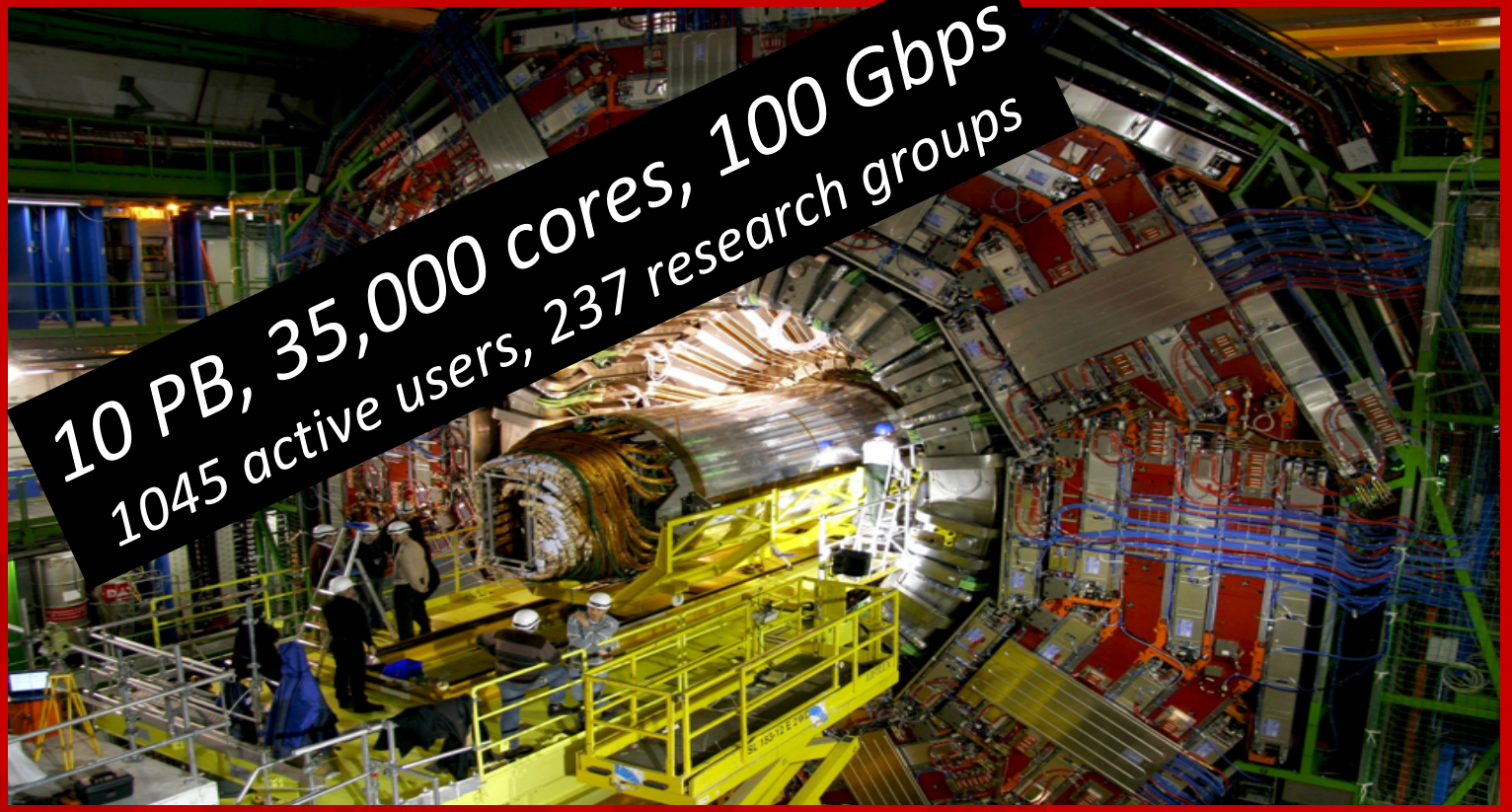
Me, My Friends and Everyone Else – Campuses and OSG

David Swanson, Director

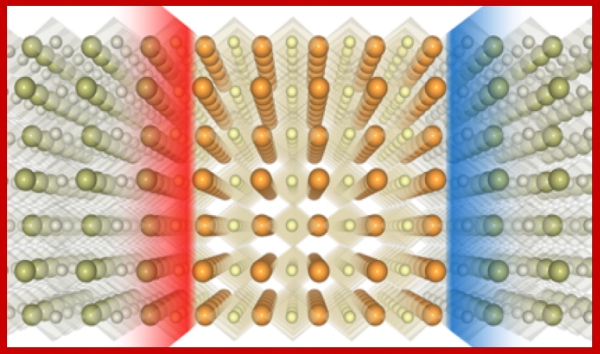
Holland Computing Center

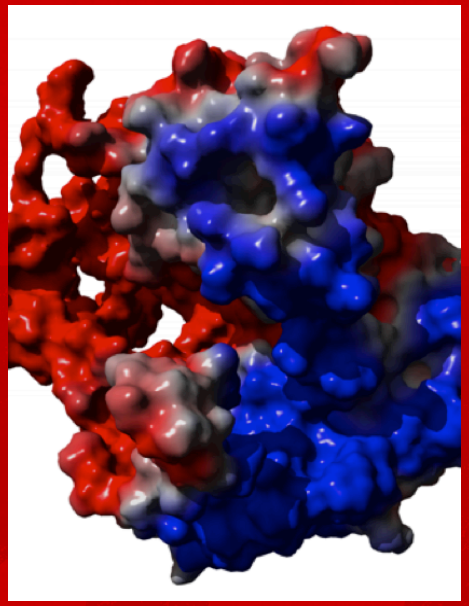
University of Nebraska-Lincoln

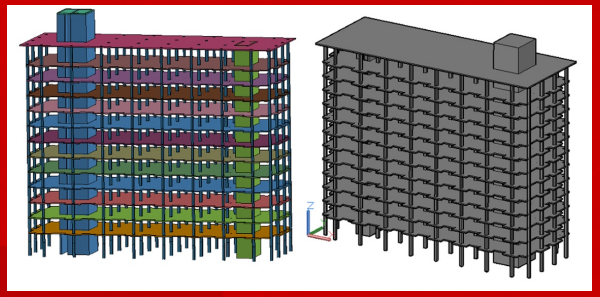
Holland Computing Center

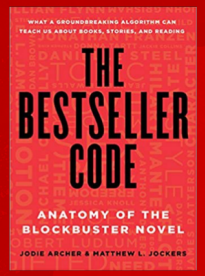


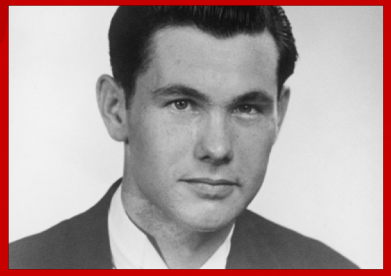
Holland Computing Center











NSF and Campus Resources

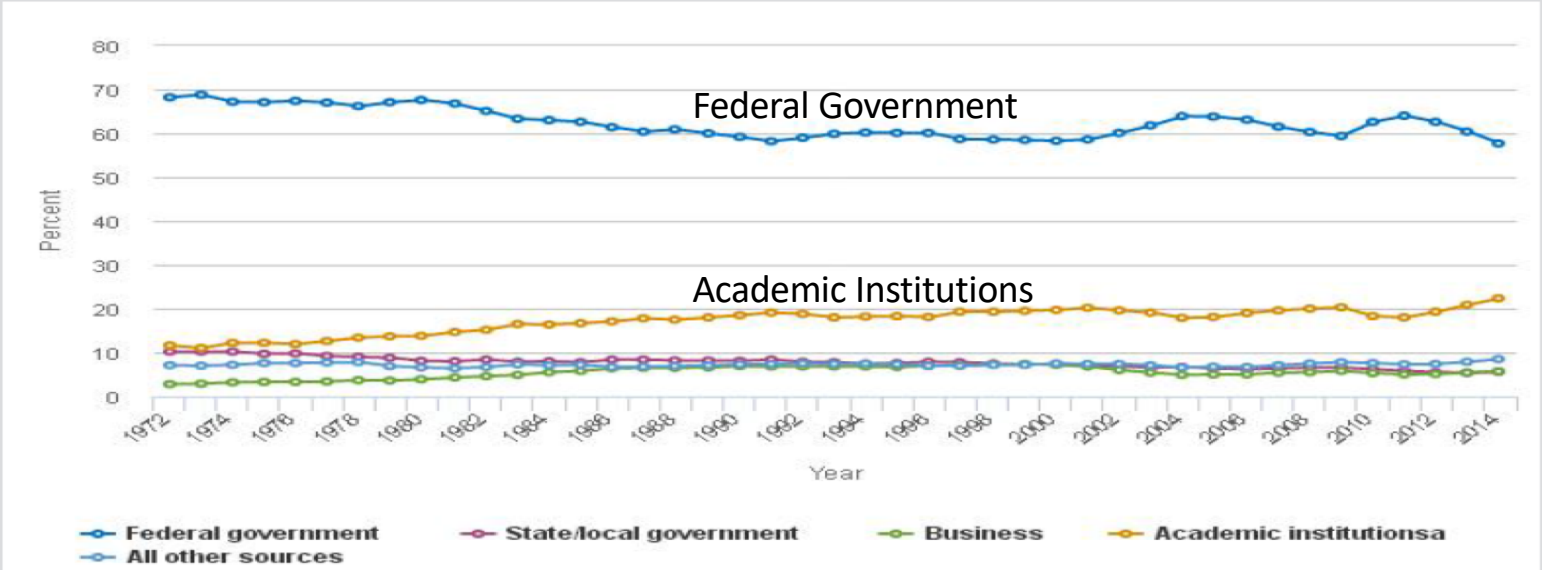
- Like other institutions across the nation, Nebraska invests into research computing to remain competitive
- From a funding perspective, the proportion spent on R & D by the Federal government with respect to Academic Institutions has decreased over the last several years
- It is critical to integrate our resources to maximize our investment
- OSG integrates HCC with more than resources, it connects us to science we'd otherwise not contribute to

Trends in Funding Academic R&D in Science and Engineering

National Science Board | **Science & Engineering Indicators 2016**

Figure 5-1

Academic S&E R&D expenditures, by source of funding: FYs 1972–2014



[a] Academic institutions' funds exclude research funds spent from multipurpose accounts.

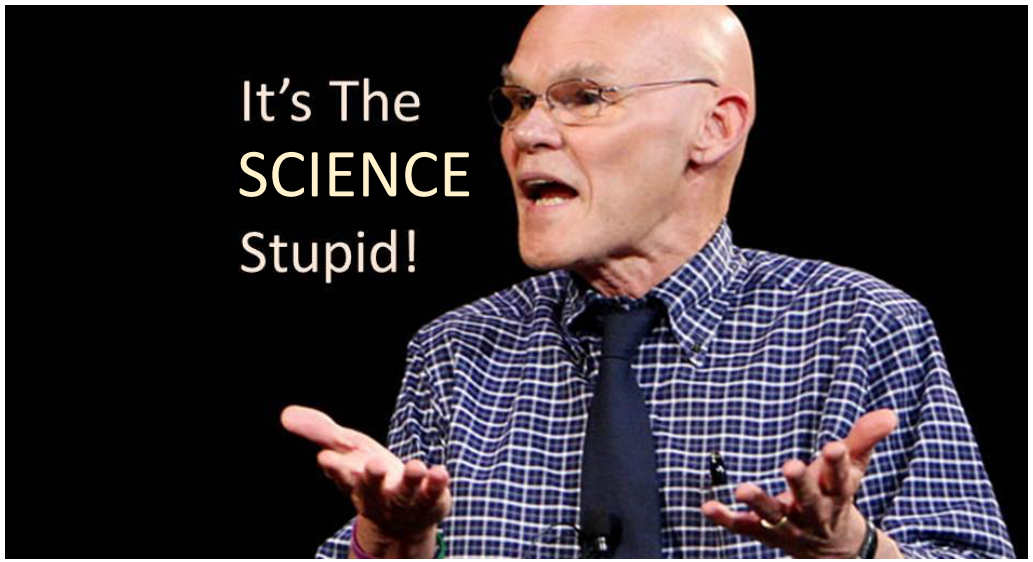
SOURCE: National Science Foundation, National Center for Science and Engineering Statistics, Higher Education Research and Development Survey.

Science and Engineering Indicators 2016

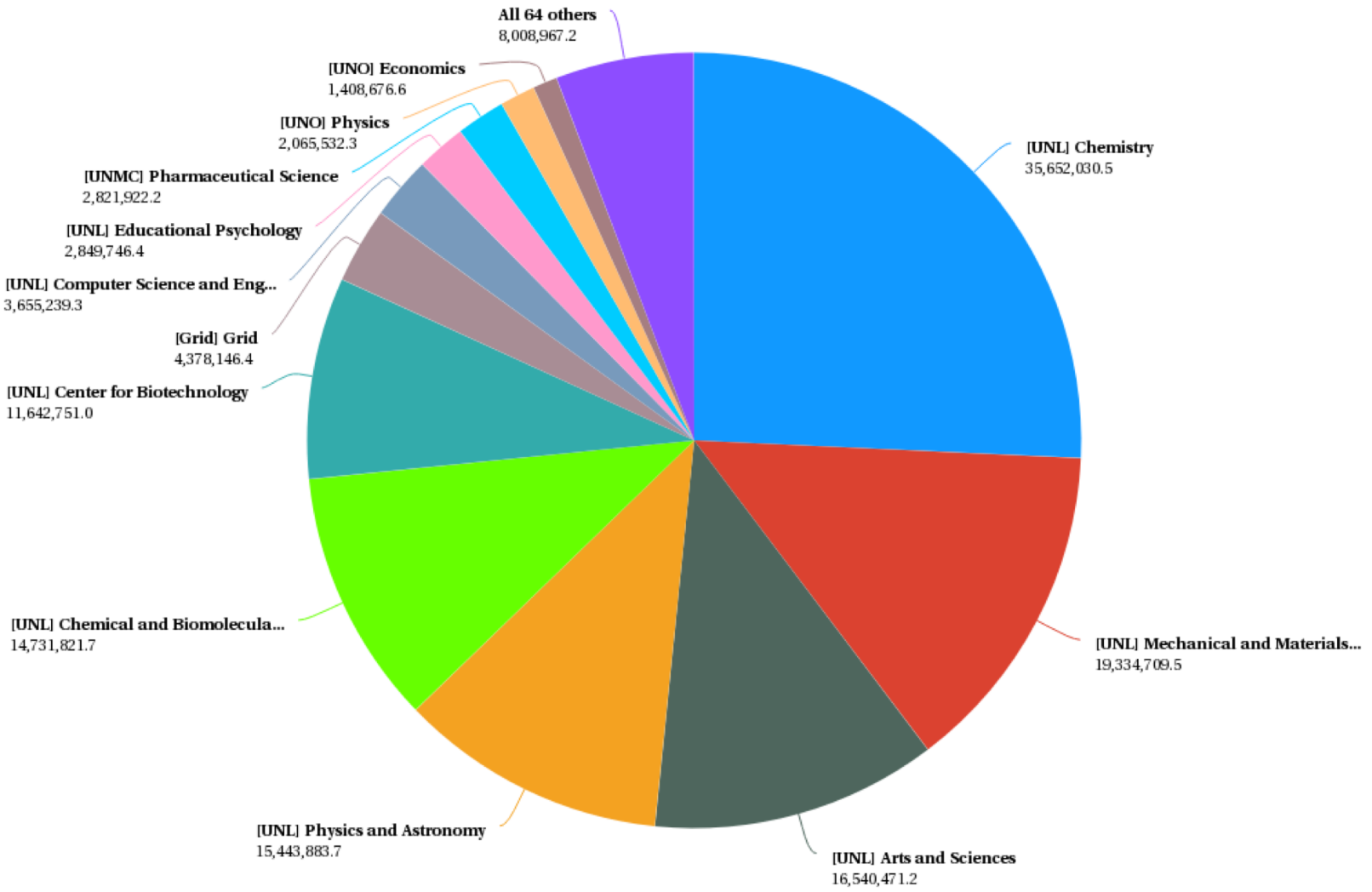
OSG appeal to Campus Resource Providers

- Better ROI for equipment
- Interact with world experts
- Improve local services
- Challenge and engage staff
- Support Science!

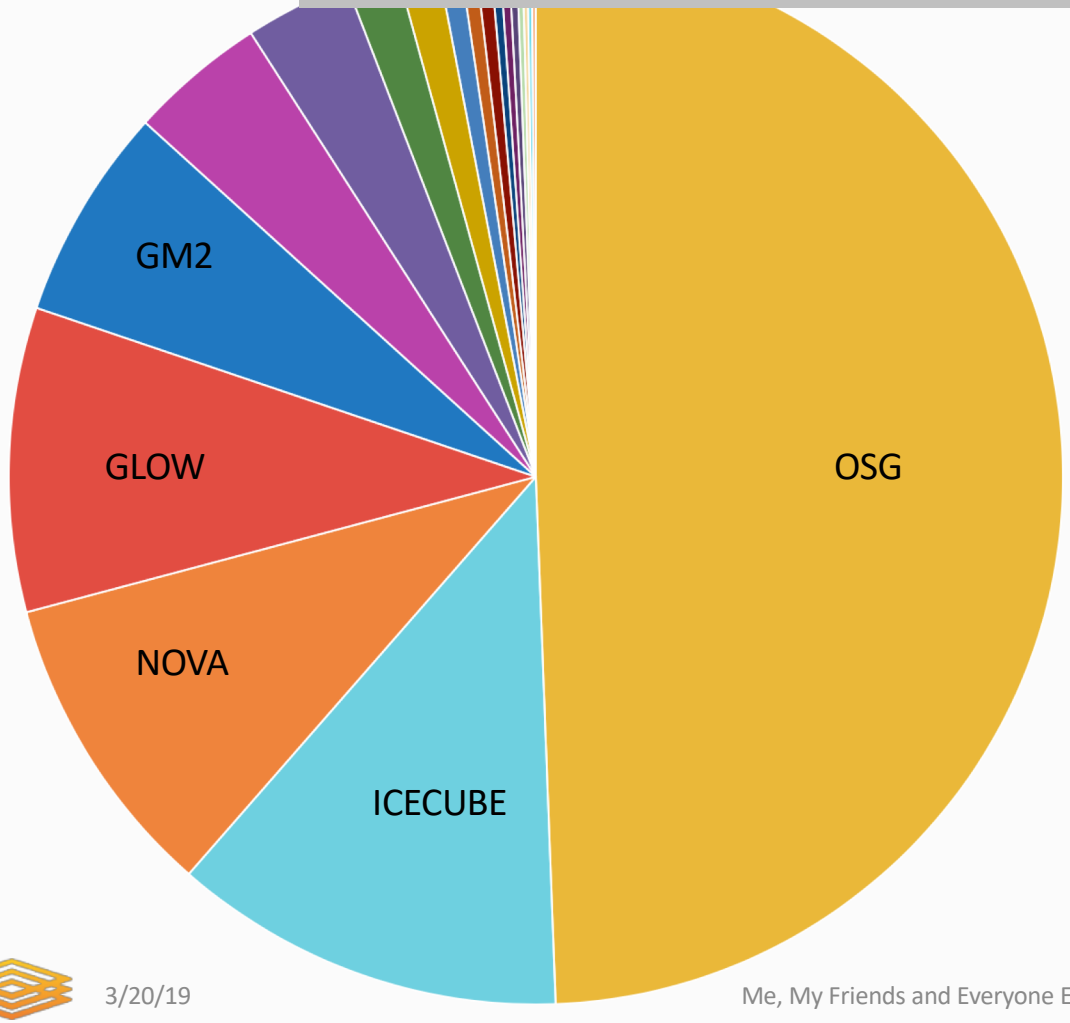
- Hosted CE's
- OSG Connect
- CC* expects it!



CPU Hours: Total: by Department

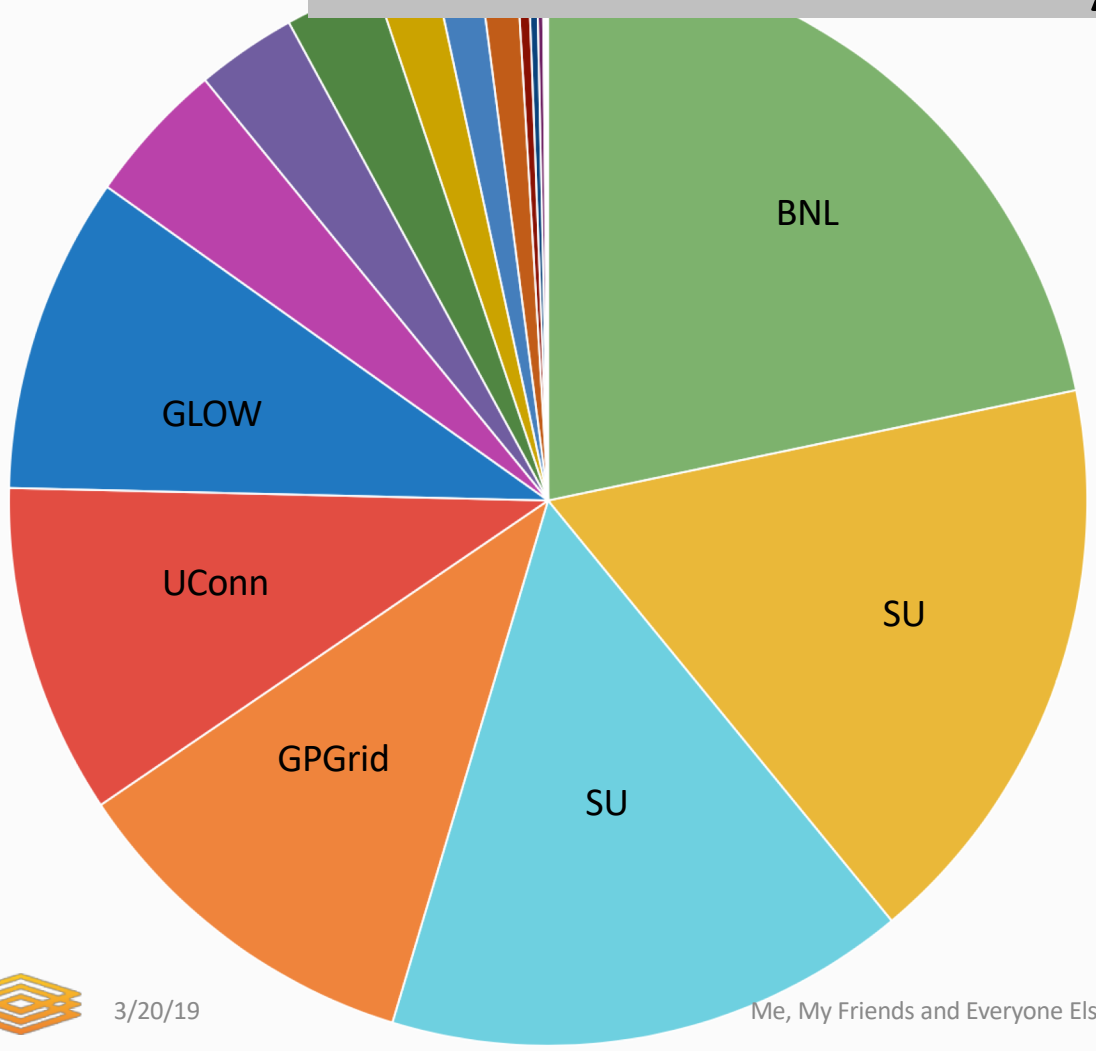


Who ran at HCC last year



	total
cms	17.2 Mil
osg	2.971 Mil
icecube	724 K
nova	567 K
glow	561 K
gm2	393 K
microboone	252 K
mu2e	195 K
/icecube/LocalGroup=tcarver	95.0 K
dune	75.1 K
seaquest	41.4 K
sbnd	28.4 K
/icecube/LocalGroup=hoshina	27.6 K
icarus	17.4 K
sbgrid	16.5 K
/gluex/LocalGroup=enp	14.53 K
minos	10.94 K
gluex	8.81 K
des	8.27 K
lariat	7.67 K

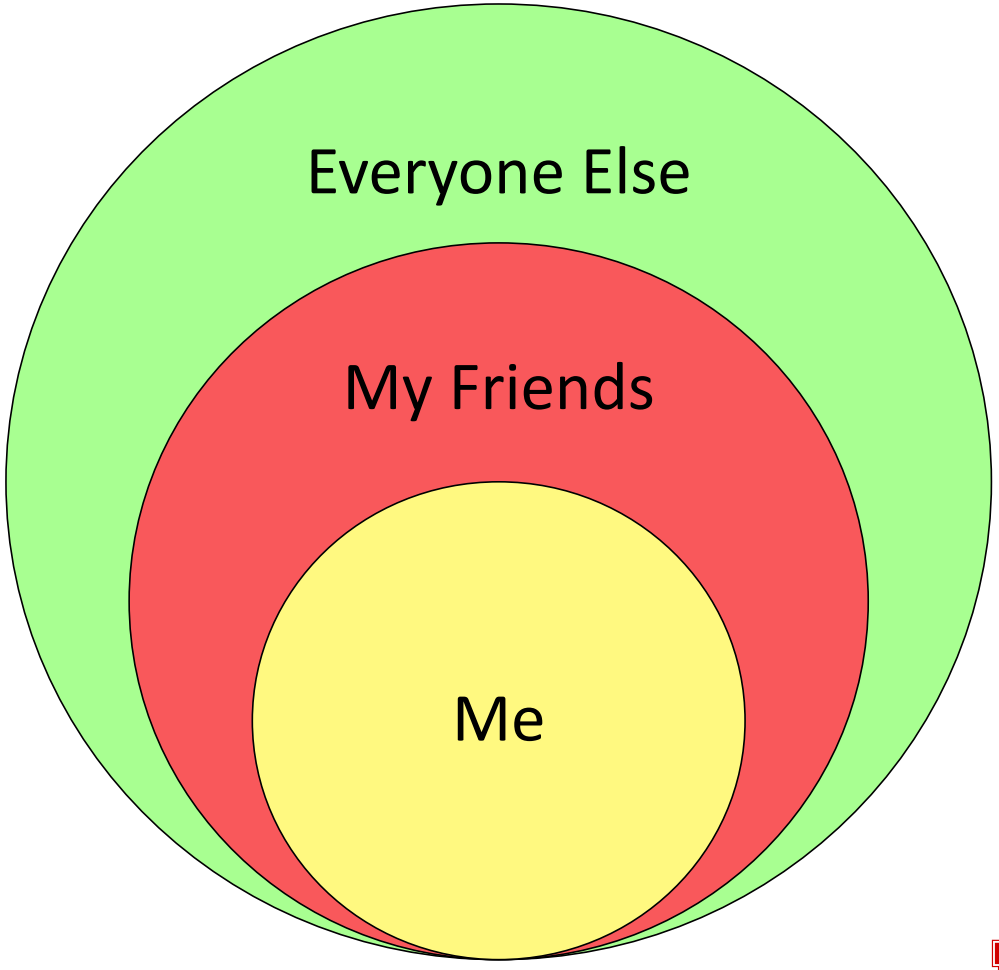
Where HCC ran last year



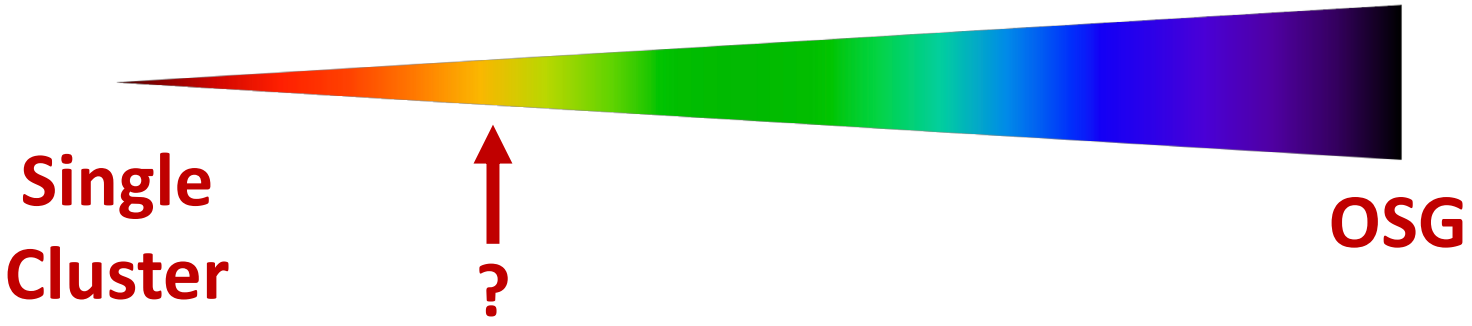
	total
BNL-ATLAS	474 K
SU-ITS-CE2	376 K
SU-ITS-CE3	341 K
GPGrid	238 K
UConn-OSG	213 K
GLOW	205 K
NUMEP_CE	93.0 K
SPRACE	65.3 K
MWT2	60.5 K
AGLT2	38.9 K
CIT_CMS_T2	29.4 K
SU-OG-CE1	24.6 K
GridUNESP_CENTRAL	7.56 K
N/A	5.63 K
Tusker	3.94 K
SU-OG-CE	1.479 K
SWT2_CPB	954
T3_US_NERSC	703
Crane	326
USCMS-FNAL-WC1	78

Spheres of resources

- Me = Local
  - Departmental or research group
- My Friends = Campus
  - Same organization
  - Colleagues
- Everyone Else = Grid
  - All affiliated resources
  - Can’t expect any favors
  - Largest scale



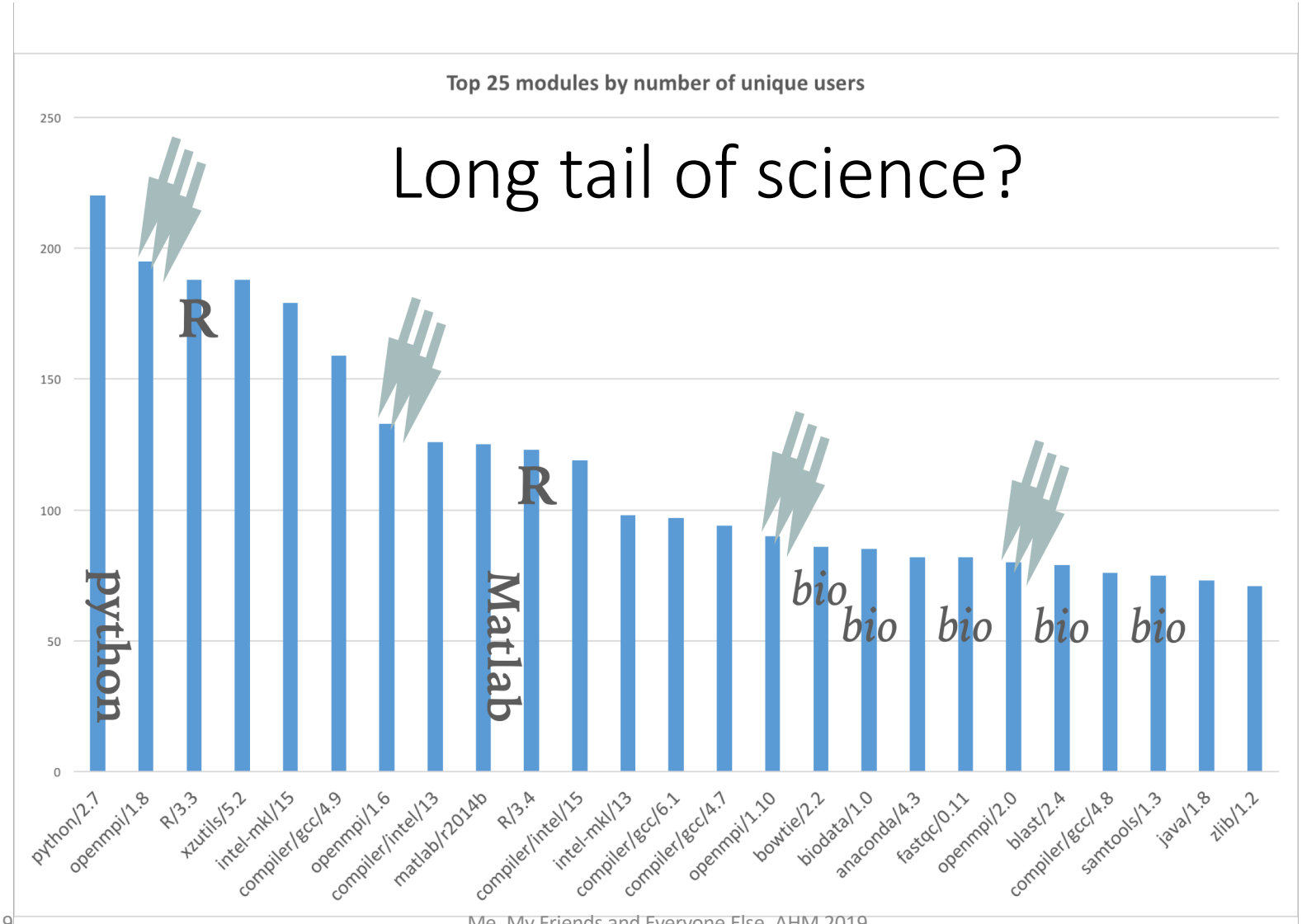
Spectrum of Collaborations



- Me, My Friends, Everyone Else
- We know how to make a single cluster and how to integrate with the OSG
- But how to enable “small” sharing?

All a resource provider wants is ...

- To know who is using resources
- To prioritize this usage (for my friends)
- To do this in a way that maintains my ownership (i.e. I want to implement a new policy on Friday afternoon unilaterally)
- To be able to combine resources in a way that reflects the social aspects of collaboration
- Science remains a team sport – but some teams are small or short-lived


Top 25 modules by number of unique users
250
200
150
100
50
0
Long tail of science?
python
R
R
Matlab
bio
bio
bio
bio
bio
python/2.7
openmpi/1.8
R/3.3
xzutils/5.2
intel-mkl/15
compiler/gcc/4.9
openmpi/1.6
compiler/intel/13
matlab/r2014b
R/3.4
compiler/intel/15
intel-mkl/13
compiler/gcc/6.1
compiler/gcc/4.7
openmpi/1.10
bowtie/2.2
biodata/1.0
anaconda/4.3
fastqc/0.11
openmpi/2.0
blast/2.4
compiler/gcc/4.8
samtools/1.3
java/1.8
zlib/1.2

In-state collaboration

- Dr. Tessa Durham Brooks
- Doane University
- Root image datasets
- Ran via HCC to OSG initially, now goes directly
- Doane accounts at HCC

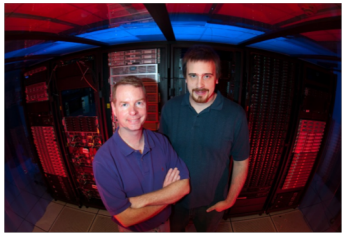
- Collaborated on CC*
- Working toward Statewide consortium
  - OneOCII

Multi-State Collaboration

Statistics, Data and Computational Science

HCC resources

Ecological Genomics



Beocat cluster

Multi-State Collaboration

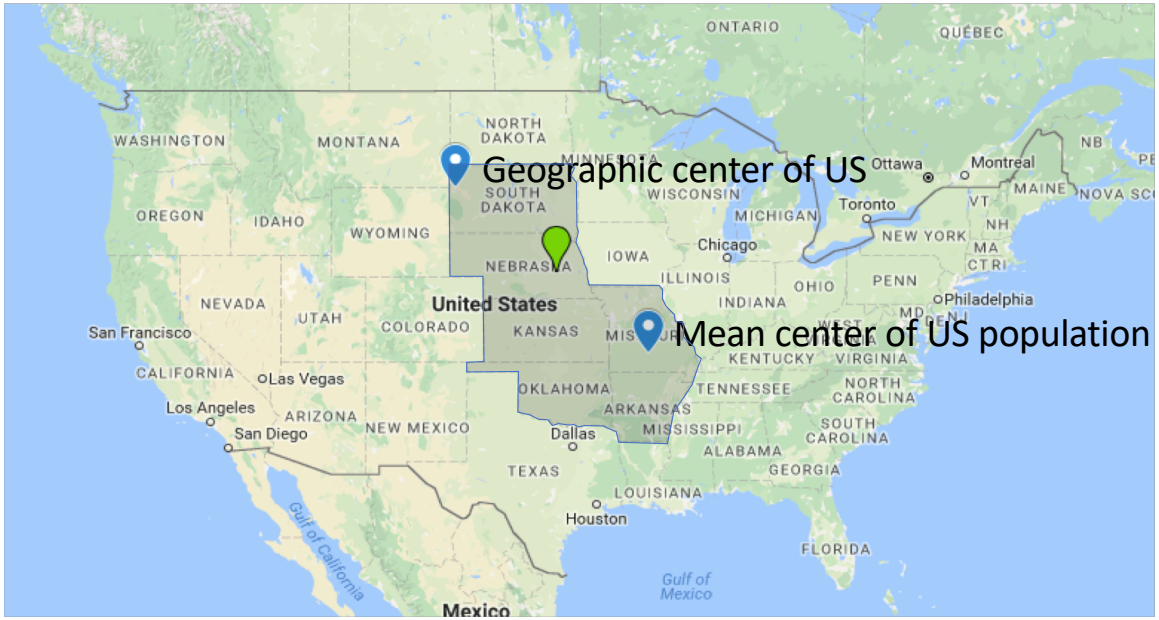
- Jennifer Clarke from Nebraska and Brad Olson from Kansas State are involved with genome to phenome research
- “Let’s collaborate and better predict crop yield!”
- EPSCoR Multi-jurisdiction award submitted
  - Critical mass across state lines
  - There’s cash, let’s be friends!
- Includes teams of students
- Resource providers from each Institution for computing, data

Multi-state Collaboration wish list

- Share data without moving it all (StashCache)
- Share processing – esp. elevated priority or reserved resources
- Familiar environment (OASIS)
- IdM handled conveniently
  - Not have to add loginID, passwd, MFA to each cluster
  - And then their students, postdocs, etc.
- Lightweight (i.e. not GUMS)
- FIONAs ftw!

Regional Organizations

- Great Plains Network
  - VO
  - FIONAs
- PRP
- RMACC
- Proximity can facilitate training, meetings
- Efficiency of scale



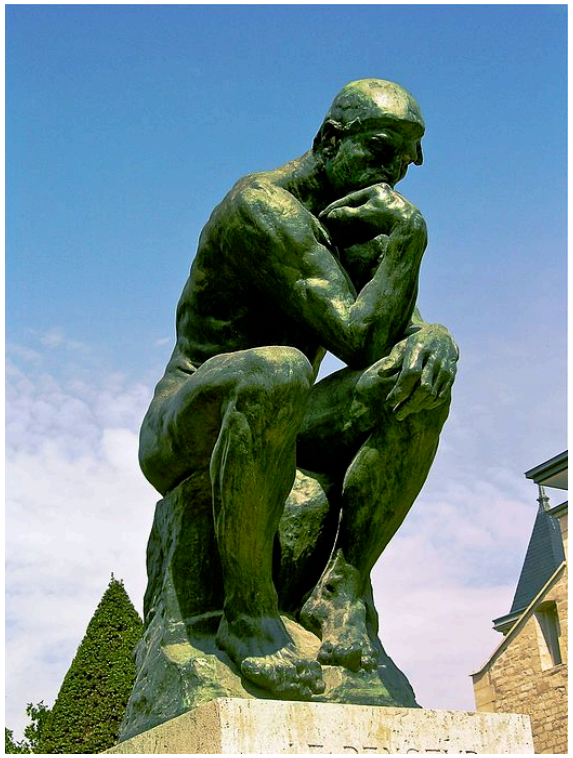
OSG developments (see other talks)

- Hosted CE / OSG-Connect
- StashCache – RO from others' storage – OSG Data Federation
- InCommon / SciTokens
- OASIS (Modules)
- Singularity
- Prioritization within Global Pool
- GRACC
- *Numerous Personnel Resources ramping up*

- Do I just want to enable … Flocking?
  - Run SLURM at both HCC, KSU
  - HCC VO could be used

Philosophical Interlude

- Technology should mirror sociology
  - Each collaborator brings some ideas and expertise to the collaboration
  - Natural if cyber-resources similarly shared
  - Currently share access to global pool of resources with fair share
  - Want that _and_ ability to aggregate and/or prioritize among a few institutions
  - Agrees with OSG principle of Local Autonomy and Control



Summary

- OSG serves both Big Science and Individual Investigators extremely well
- Small, ad-hoc groups and long tail users are a “My Friends” scenario
- OSG has intentionally made great strides in both support and user features for long tail science
- Resource providers would find it attractive to have more fine-grained control readily available