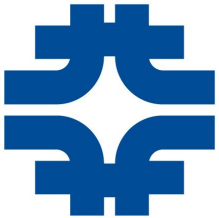
Diversity of Science on the OSG

Open Science Grid

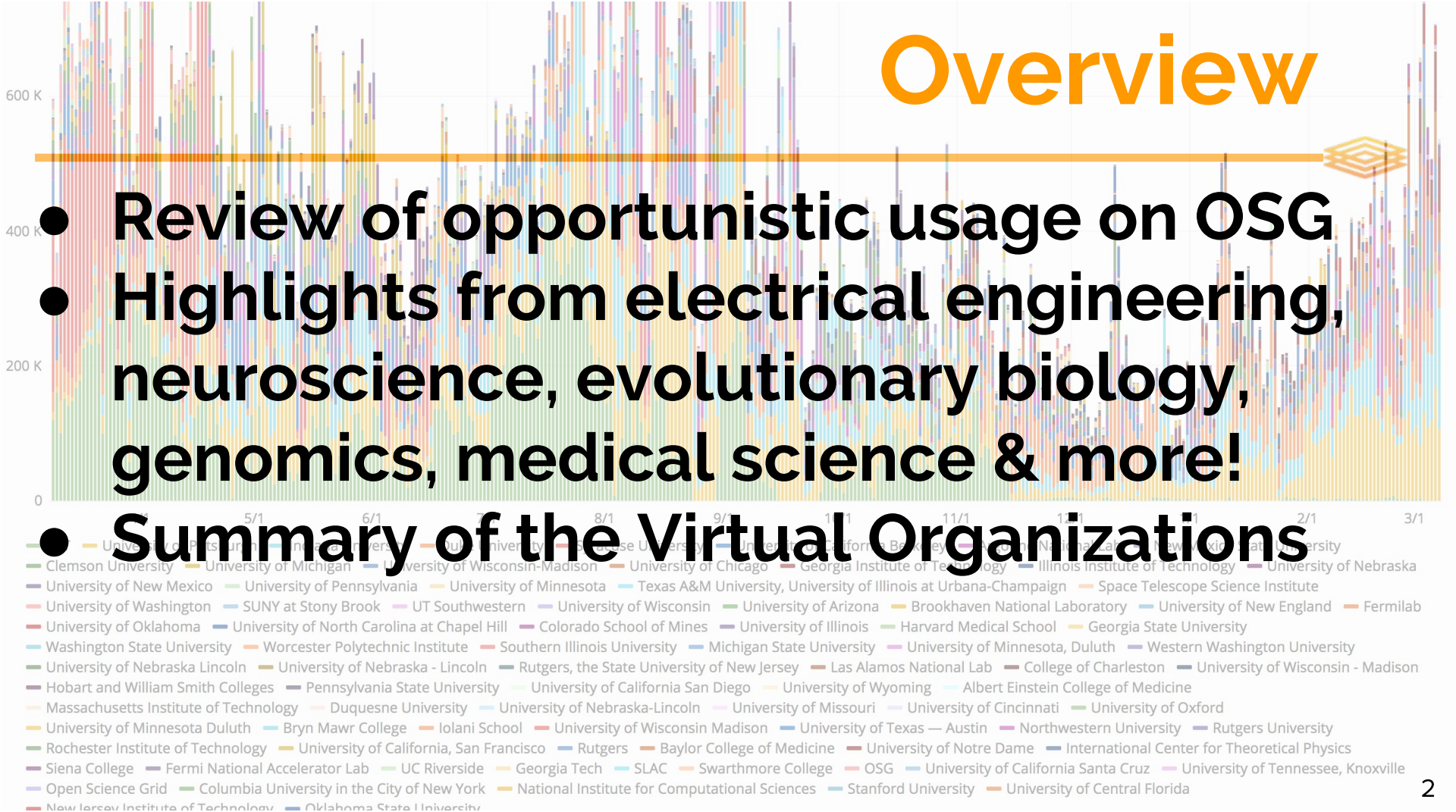
Rob Gardner • University of Chicago
Ken Herner • Fermilab



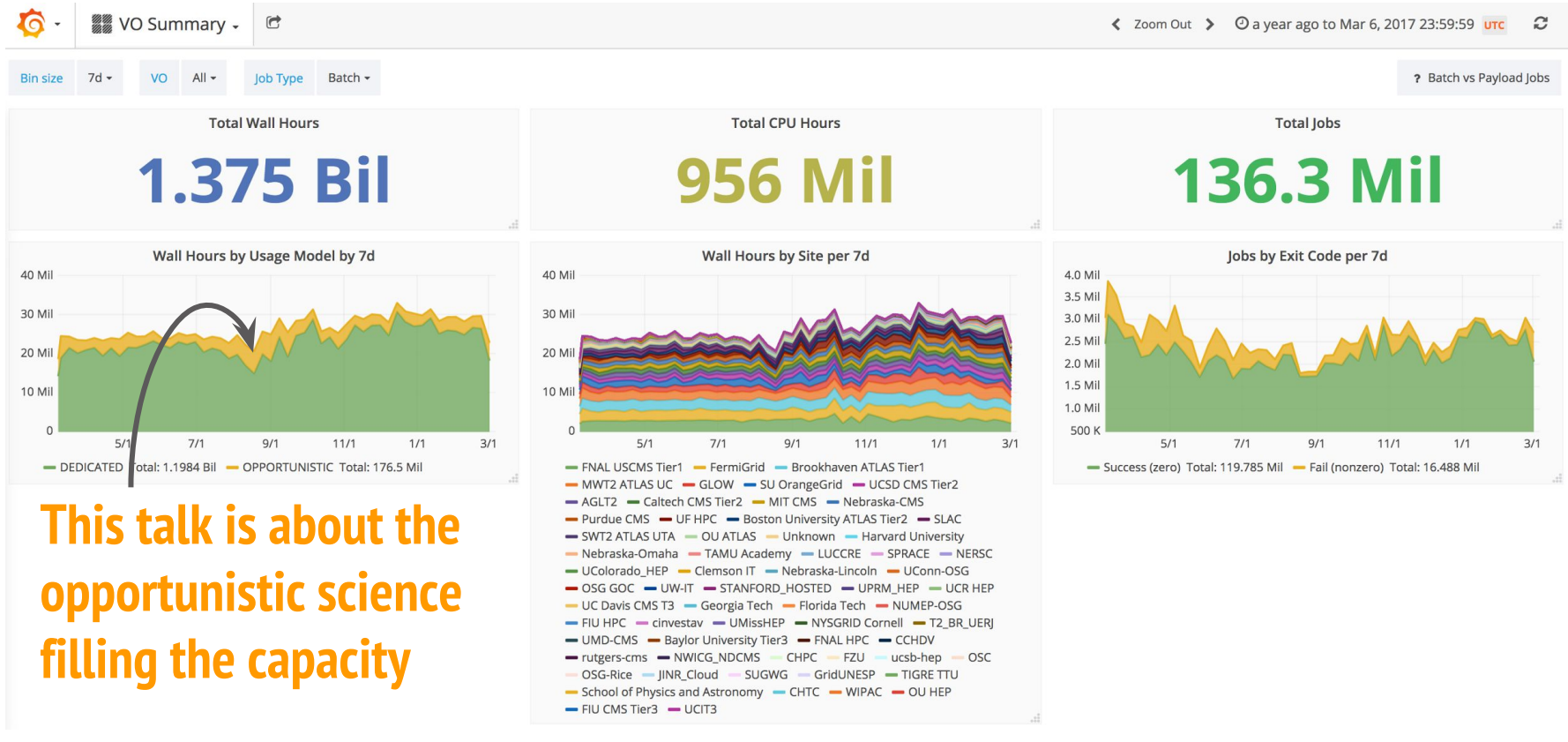
OSG All Hands Meeting, University of California San Diego, March 6-9, 2017

Overview

- Review of opportunistic usage on OSG
- Highlights from electrical engineering, neuroscience, evolutionary biology, genomics, medical science & more!
- Summary of the Virtual Organizations



Total Production (VO Summary)



This talk is about the opportunistic science filling the capacity

VO Summary w/o ATLAS and CMS

Total Wall Hours

390 Mil

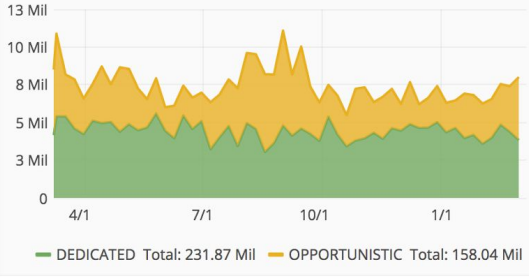
Total CPU Hours

263 Mil

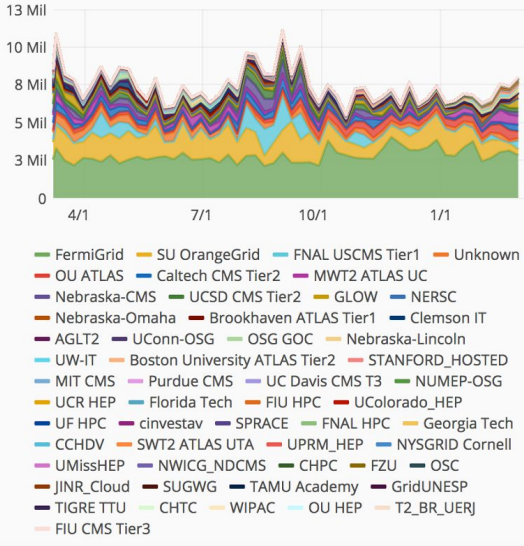
Total Jobs

31.3 Mil

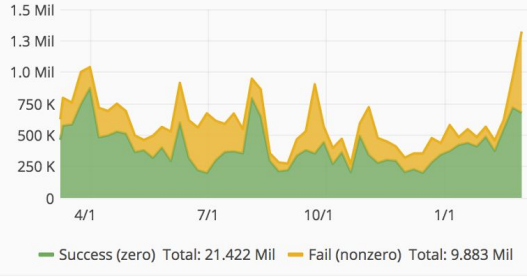
Wall Hours by Usage Model by 7d



Wall Hours by Site per 7d



Jobs by Exit Code per 7d



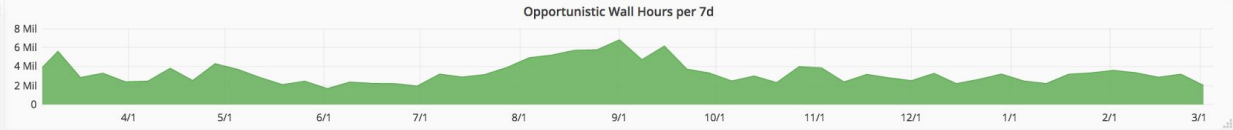
About 28% of batch hours in the past year were not ATLAS or CMS

Opportunistic

Opportunistic Usage

Zoom Out | Last 1 year UTC

Bin auto | Filter +

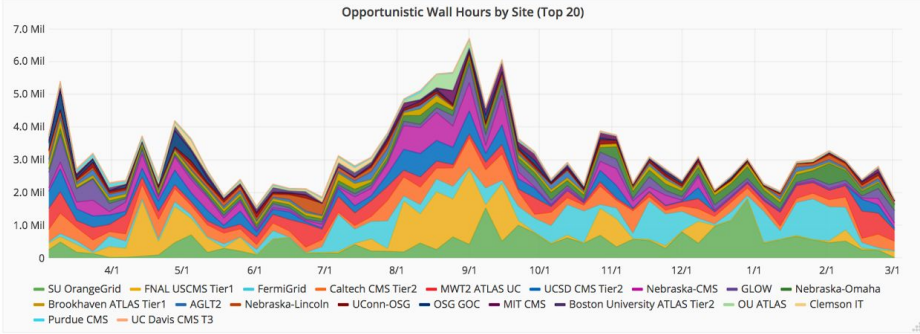


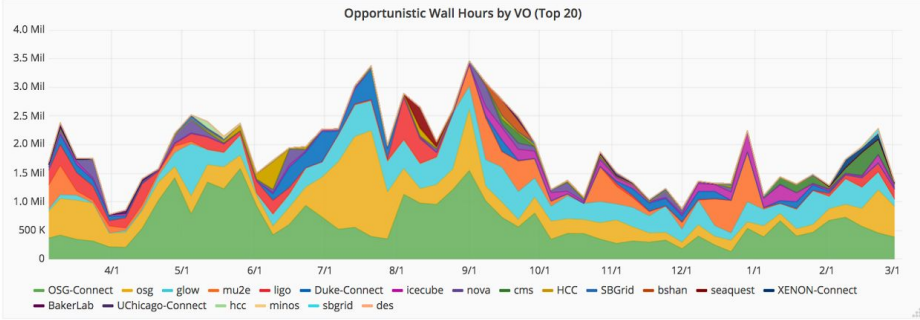
Total Opportunistic Wall Hours

176 Mil

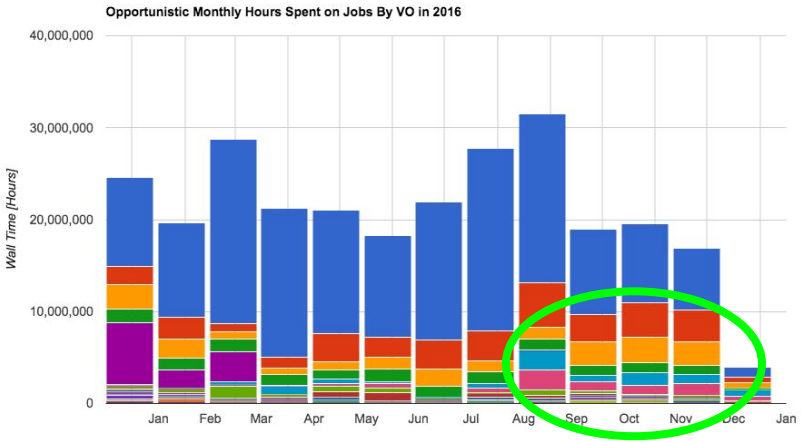
Top 20 Sites Providing Opportunistic Usage

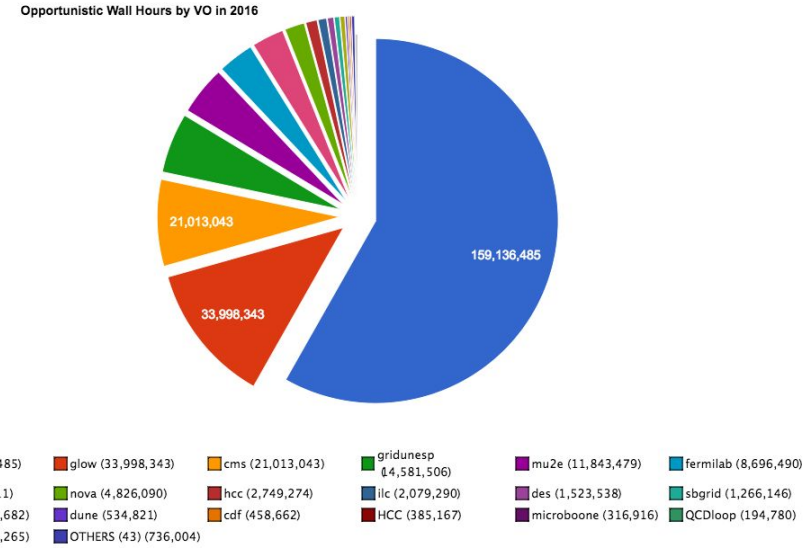
OIM_Site	OIM_Facility	Wall Hours
SU OrangeGrid	Syracuse University	24 Mil
FNAL USCMS Tier1	Fermi National Accelerator Laboratory	19 Mil
FermiGrid	Fermi National Accelerator Laboratory	19 Mil
Caltech CMS Tier2	California Institute of Technology	17 Mil
MWT2 ATLAS UC	University of Chicago	15 Mil
UCSD CMS Tier2	University of California San Diego	13 Mil
Nebraska-CMS	University of Nebraska	13 Mil
GLOW	University of Wisconsin	10 Mil
Nebraska-Omaha	University of Nebraska	9 Mil
Brookhaven ATLAS Tier1	Brookhaven National Laboratory	5 Mil
AGLT2	University of Michigan	4 Mil
Nebraska-Lincoln	University of Nebraska	4 Mil
UConn-OSG	University of Connecticut	3 Mil
OSG GOC	Indiana University	3 Mil
MIT CMS	Massachusetts Institute of Technology	3 Mil
Boston University ATLAS Tier2	Boston University	2 Mil
OU ATLAS	University of Oklahoma	1 Mil
Clemson IT	Clemson University	1 Mil
Purdue CMS	Purdue University	1 Mil
UC Davis CMS T3	University of California Davis	1 Mil





Opportunistic VO usage trends in 2016



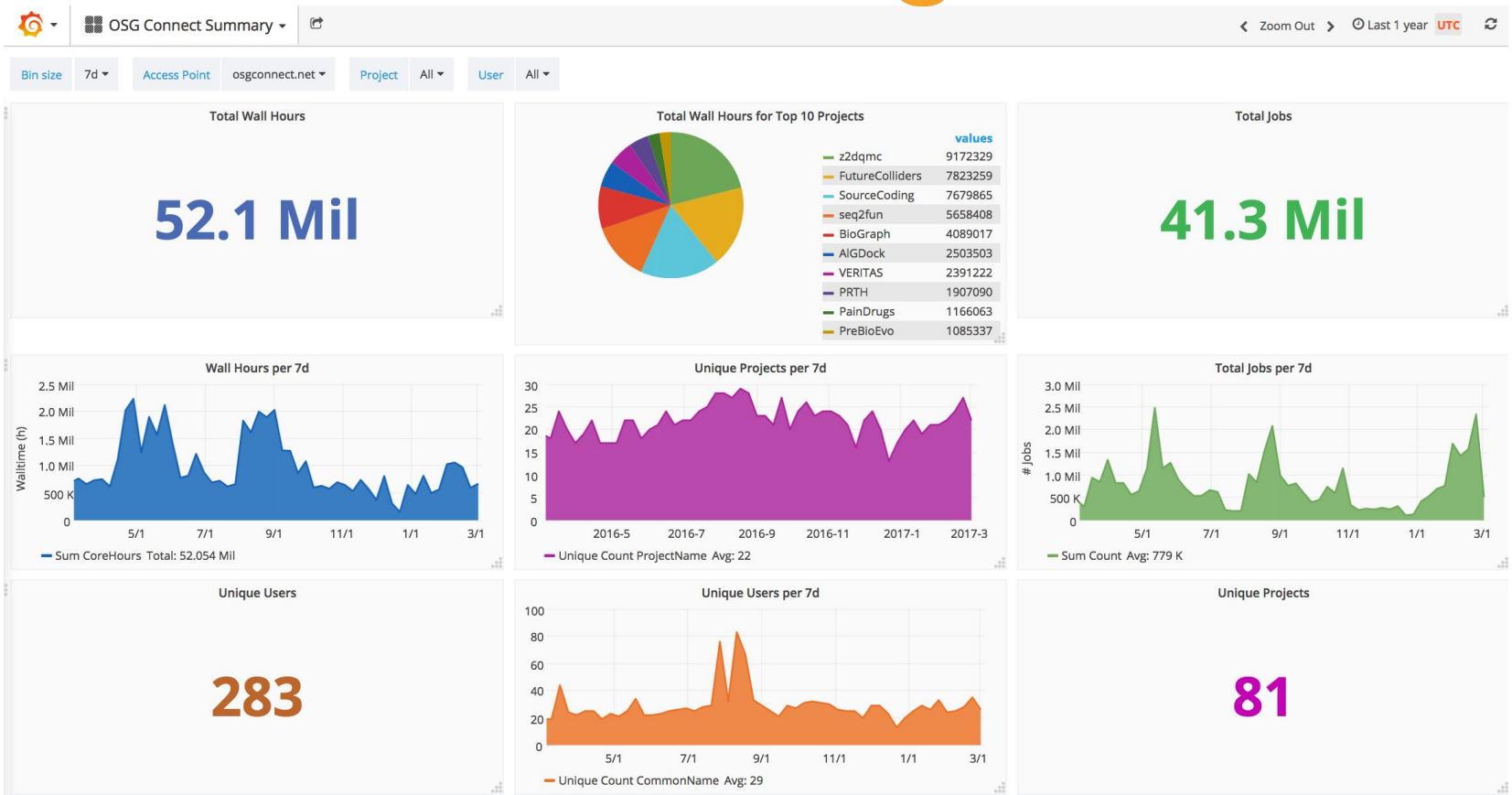


N.B. Mu2e folded into Fermilab in March

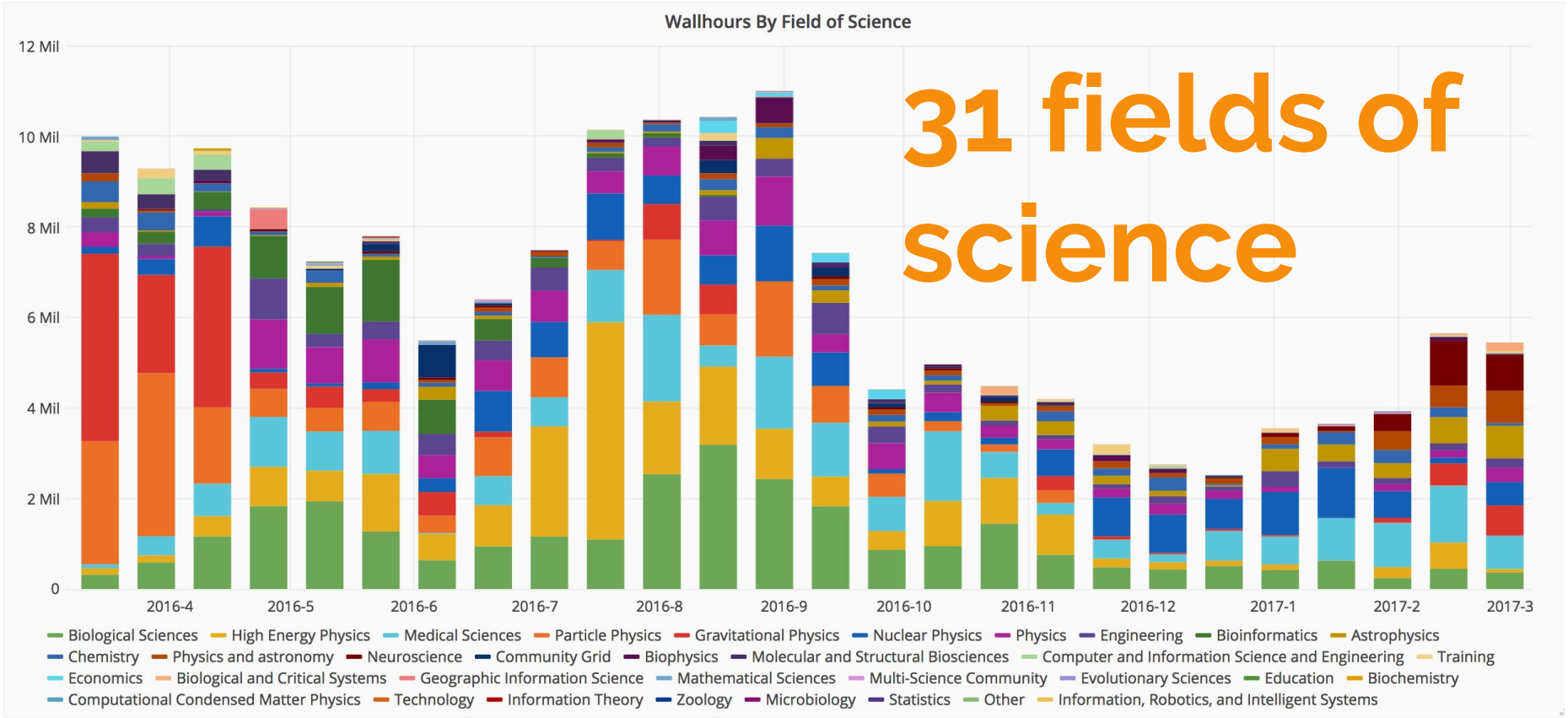
Large ATLAS and CMS increases late in year:

Expected to continue

osg connect

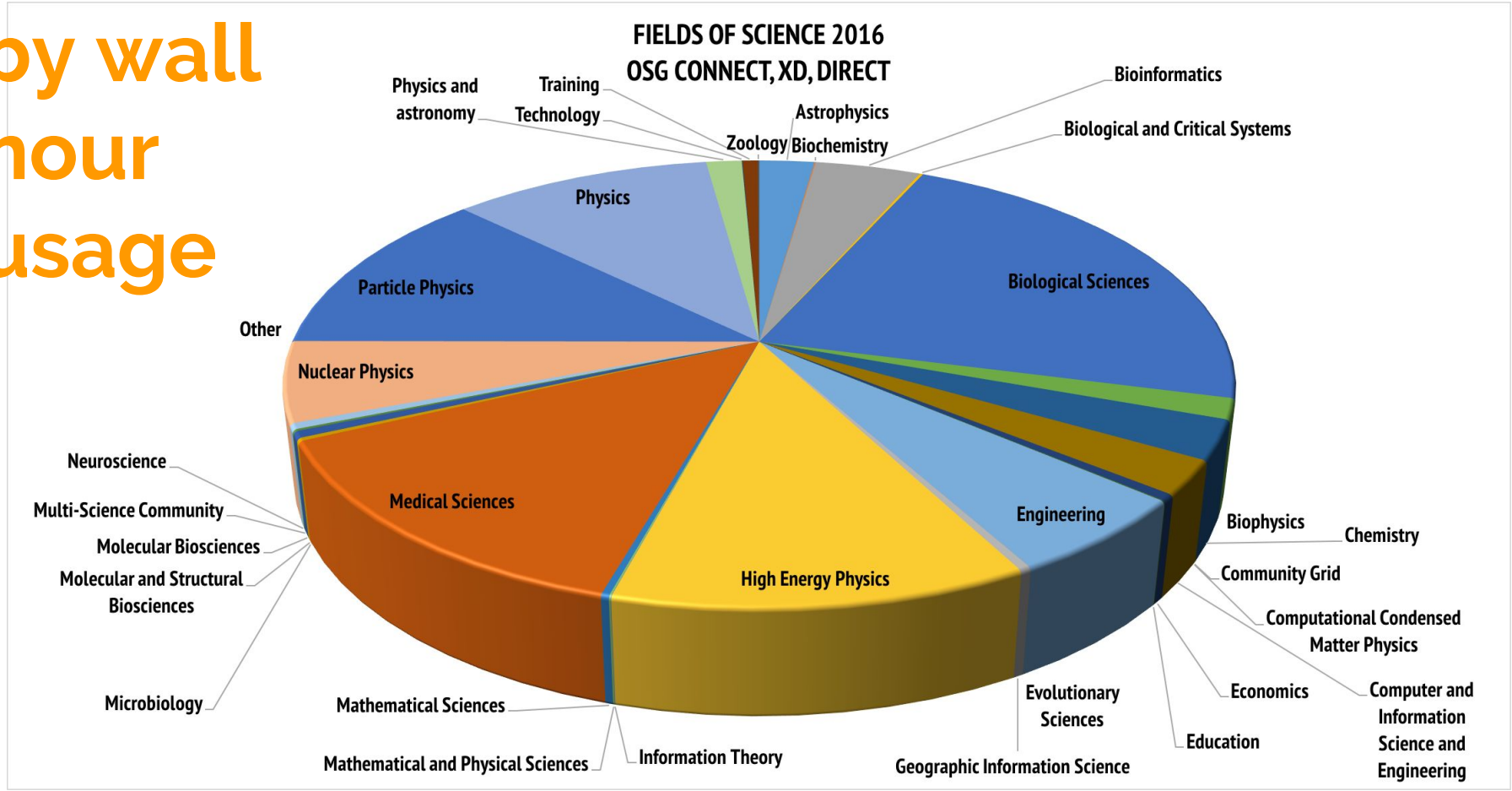

OSG Connect Summary
Zoom Out
Last 1 year
UTC
Bin size
7d
Access Point
osgconnect.net
Project
All
User
All
Total Wall Hours
52.1 Mil
Total Wall Hours for Top 10 Projects
values
z2dqmc 9172329
FutureColliders 7823259
SourceCoding 7679865
seq2fun 5658408
BioGraph 4089017
AlGDock 2503503
VERITAS 2391222
PRTH 1907090
PainDrugs 1166063
PreBioEvo 1085337
Total Jobs
41.3 Mil
Wall Hours per 7d
Walltime (h)
2.5 Mil
2.0 Mil
1.5 Mil
1.0 Mil
500 K
0
5/1
7/1
9/1
11/1
1/1
3/1
Sum CoreHours Total: 52.054 Mil
Unique Projects per 7d
30
25
20
15
10
5
0
2016-5
2016-7
2016-9
2016-11
2017-1
2017-3
Unique Count ProjectName Avg: 22
Total Jobs per 7d
Jobs
3.0 Mil
2.5 Mil
2.0 Mil
1.5 Mil
1.0 Mil
500 K
0
5/1
7/1
9/1
11/1
1/1
3/1
Sum Count Avg: 779 K
Unique Users
283
Unique Users per 7d
100
80
60
40
20
0
5/1
7/1
9/1
11/1
1/1
3/1
Unique Count CommonName Avg: 29
Unique Projects
81

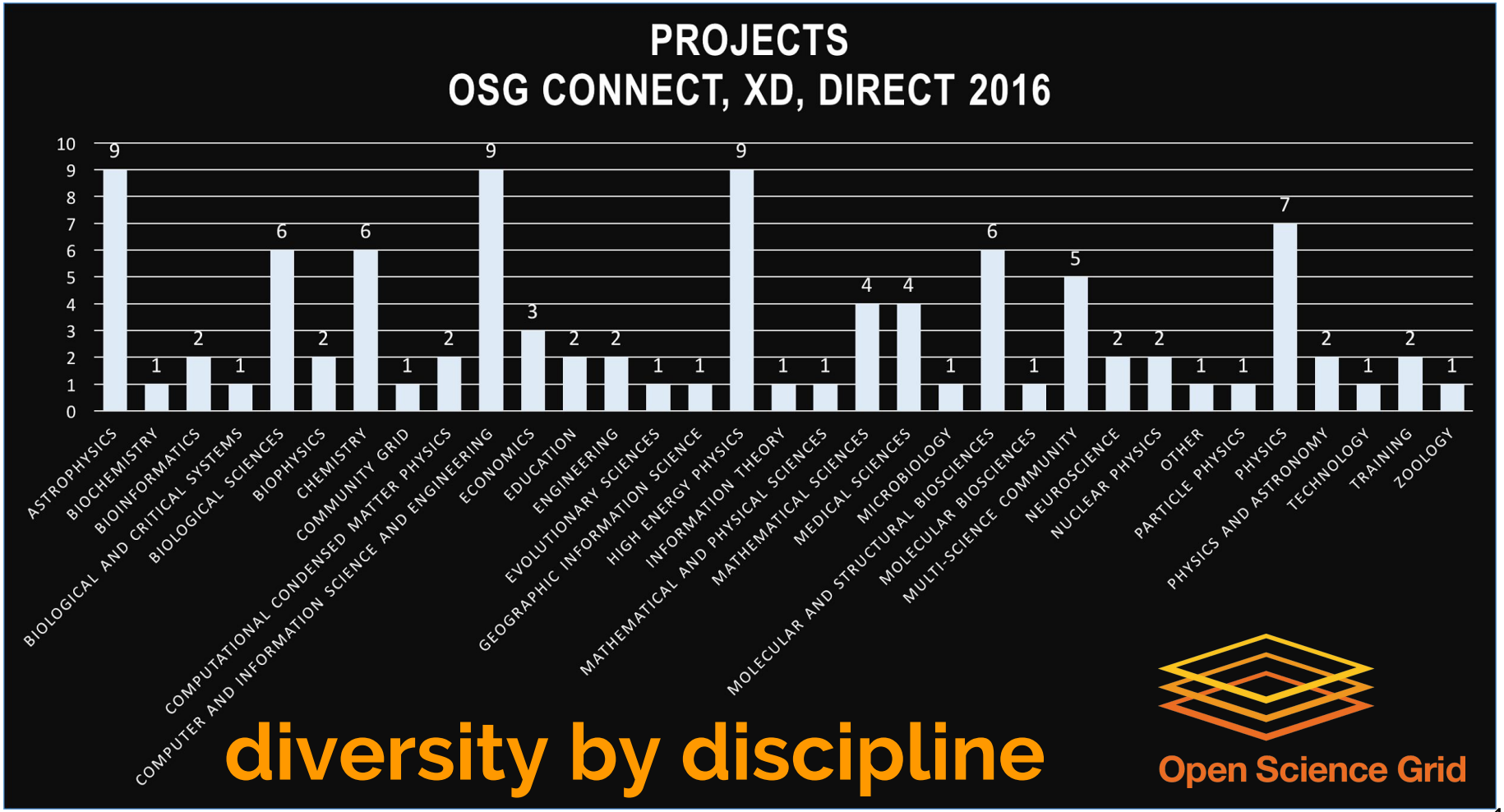
By Field Of Science

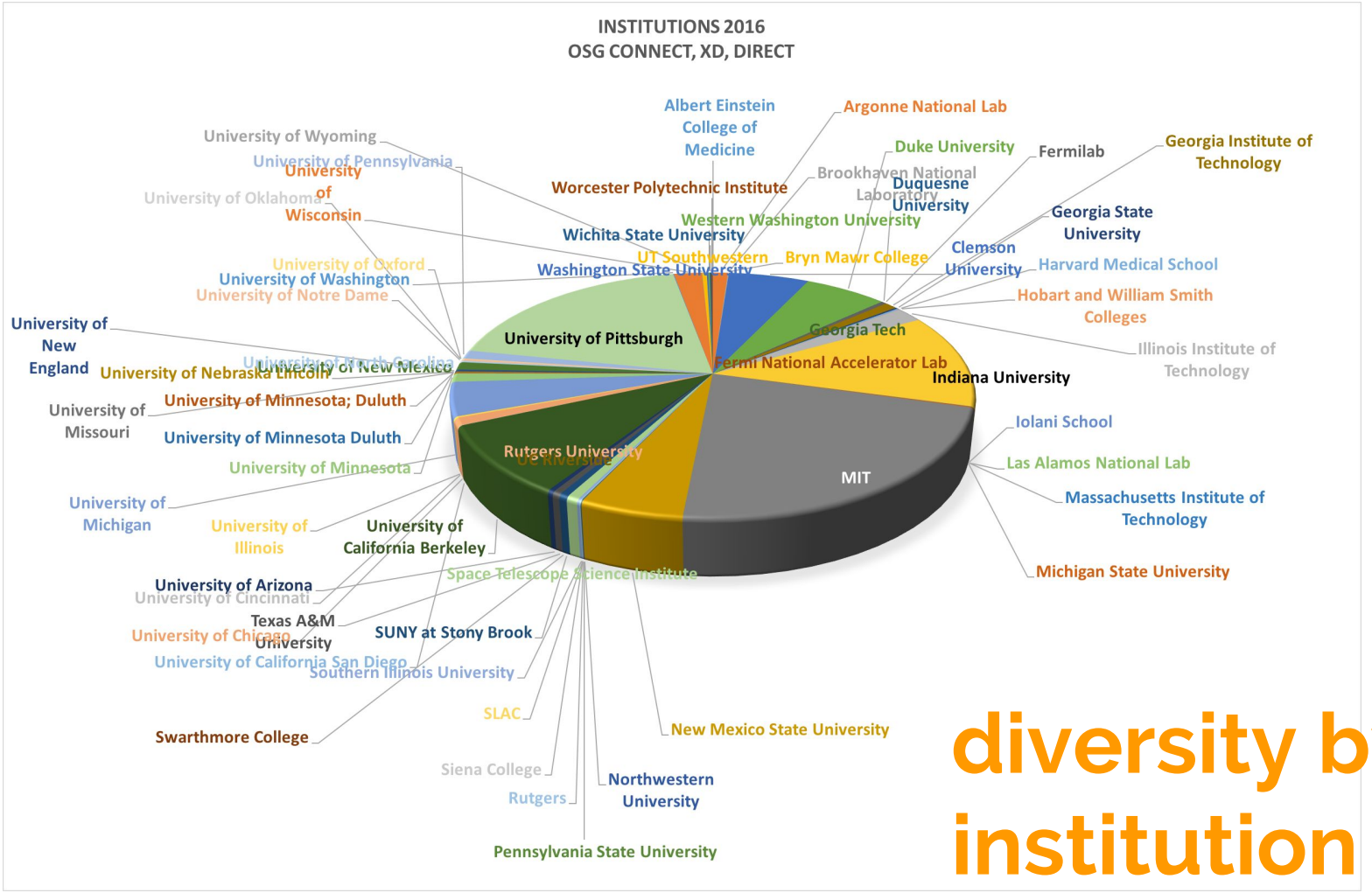


31 fields of science

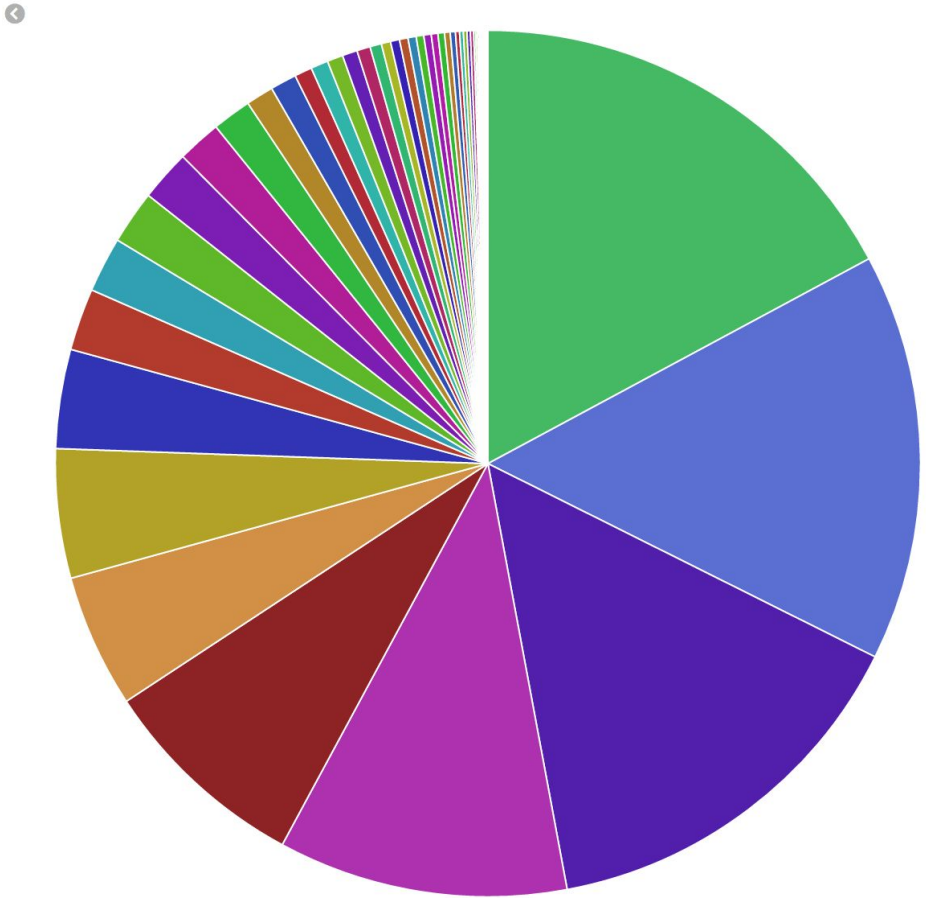
by wall hour usage



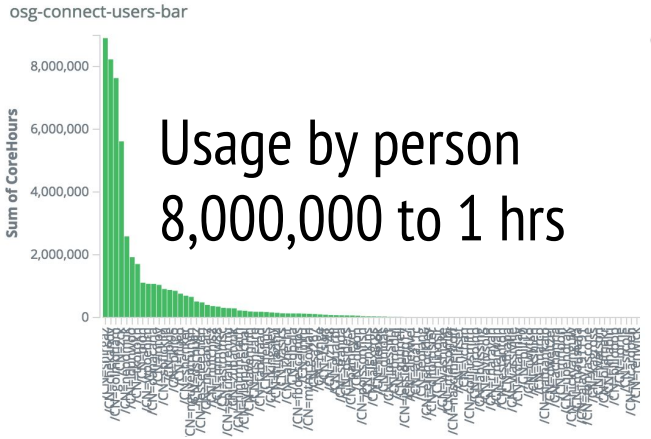

PROJECTS
OSG CONNECT, XD, DIRECT 2016
10
9
8
7
6
5
4
3
2
1
0
9
1
2
1
6
2
6
1
2
9
3
2
2
1
1
9
1
1
4
4
1
6
1
5
2
2
1
1
7
2
1
2
1
ASTROPHYSICS
BIOCHEMISTRY
BIOINFORMATICS
BIOLOGICAL AND CRITICAL SYSTEMS
BIOLOGICAL SCIENCES
BIOPHYSICS
CHEMISTRY
COMMUNITY GRID
COMPUTATIONAL CONDENSED MATTER PHYSICS
COMPUTER AND INFORMATION SCIENCE AND ENGINEERING
ECONOMICS
EDUCATION
ENGINEERING
EVOLUTIONARY SCIENCES
GEOGRAPHIC INFORMATION SCIENCE
HIGH ENERGY PHYSICS
INFORMATION THEORY
MATHEMATICAL AND PHYSICAL SCIENCES
MATHEMATICAL SCIENCES
MEDICAL SCIENCES
MICROBIOLOGY
MOLECULAR AND STRUCTURAL BIOSCIENCES
MOLECULAR BIOSCIENCES
MULTI-SCIENCE COMMUNITY
NEUROSCIENCE
NUCLEAR PHYSICS
OTHER
PARTICLE PHYSICS
PHYSICS
PHYSICS AND ASTRONOMY
TECHNOLOGY
TRAINING
ZOOLOGY
diversity by discipline
Open Science Grid



diversity by institution



- z2dqmc
- FutureColliders
- SourceCoding
- seq2fun
- BioGraph
- VERITAS
- AlGDock
- PRTH
- PainDrugs
- EvolSims
- PreBioEvo
- ConnectTrain
- numfpi
- DemandSC
- EvoProtDrug
- IceCube
- EIC
- CombinedPS
- nicesims
- molcryst
- Perchlorate
- NSNM
- SysBioEdu
- microphases
- atlas-org-uchicago
- fluidsim
- LiuLab
- uchicago
- RicePhenomics
- SNOplus
- duke-boolnet
- OSG-Staff
- freesurfer
- GeoTunnel
- holosim
- UserSchool2015
- SBGrid
- NSLS2ID
- ContinuousIntegration
- EvolvingAI
- TDAstats
- N/A
- MS-EinDRC
- SouthPoleTelescope
- mab
- duke-CMT
- xenon1t
- Unknown
- Paniceae-trans
- PathSpaceHMC



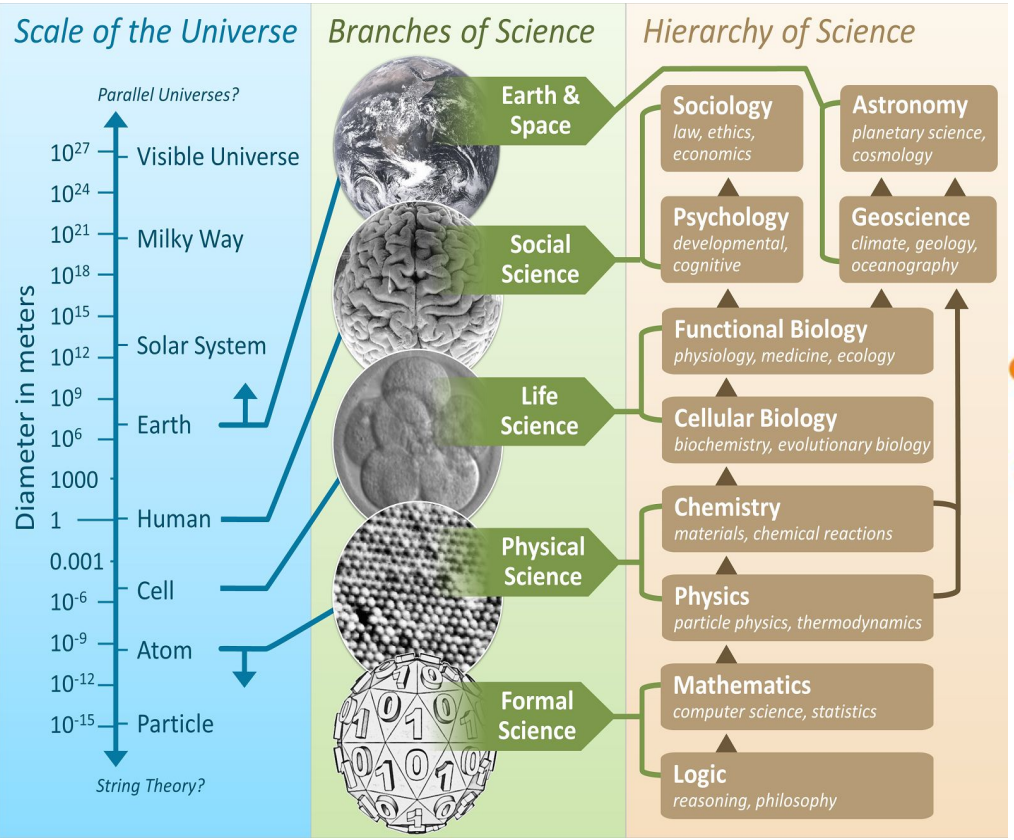
Usage by person
8,000,000 to 1 hrs

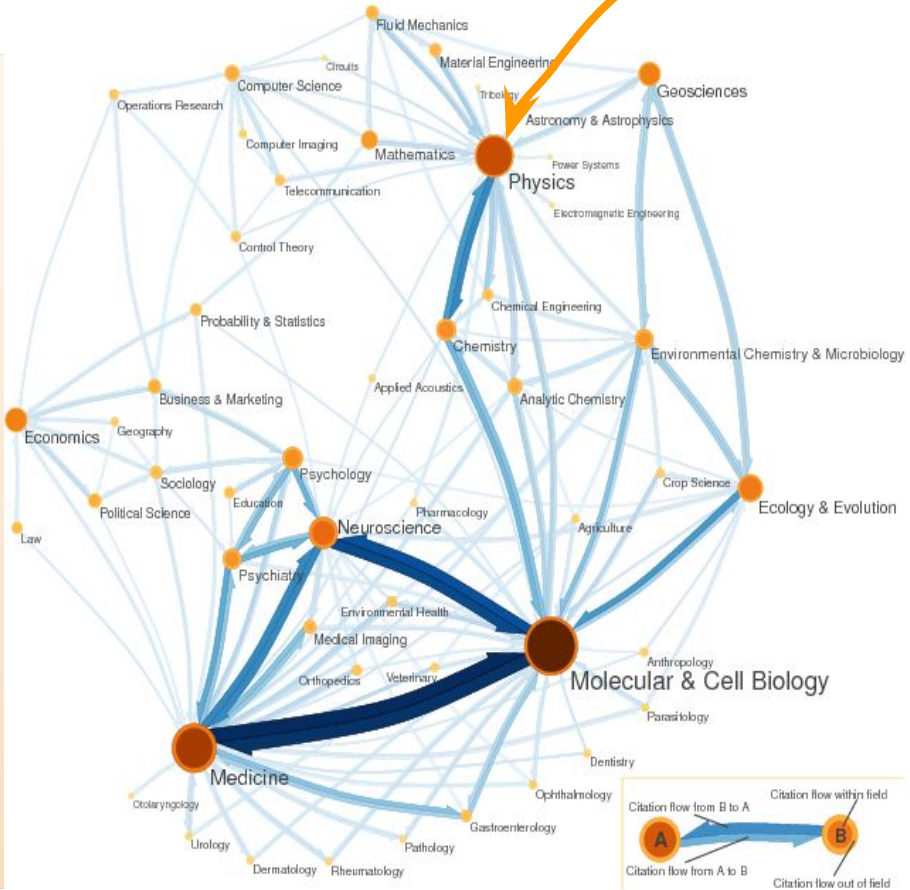
diversity by job scale

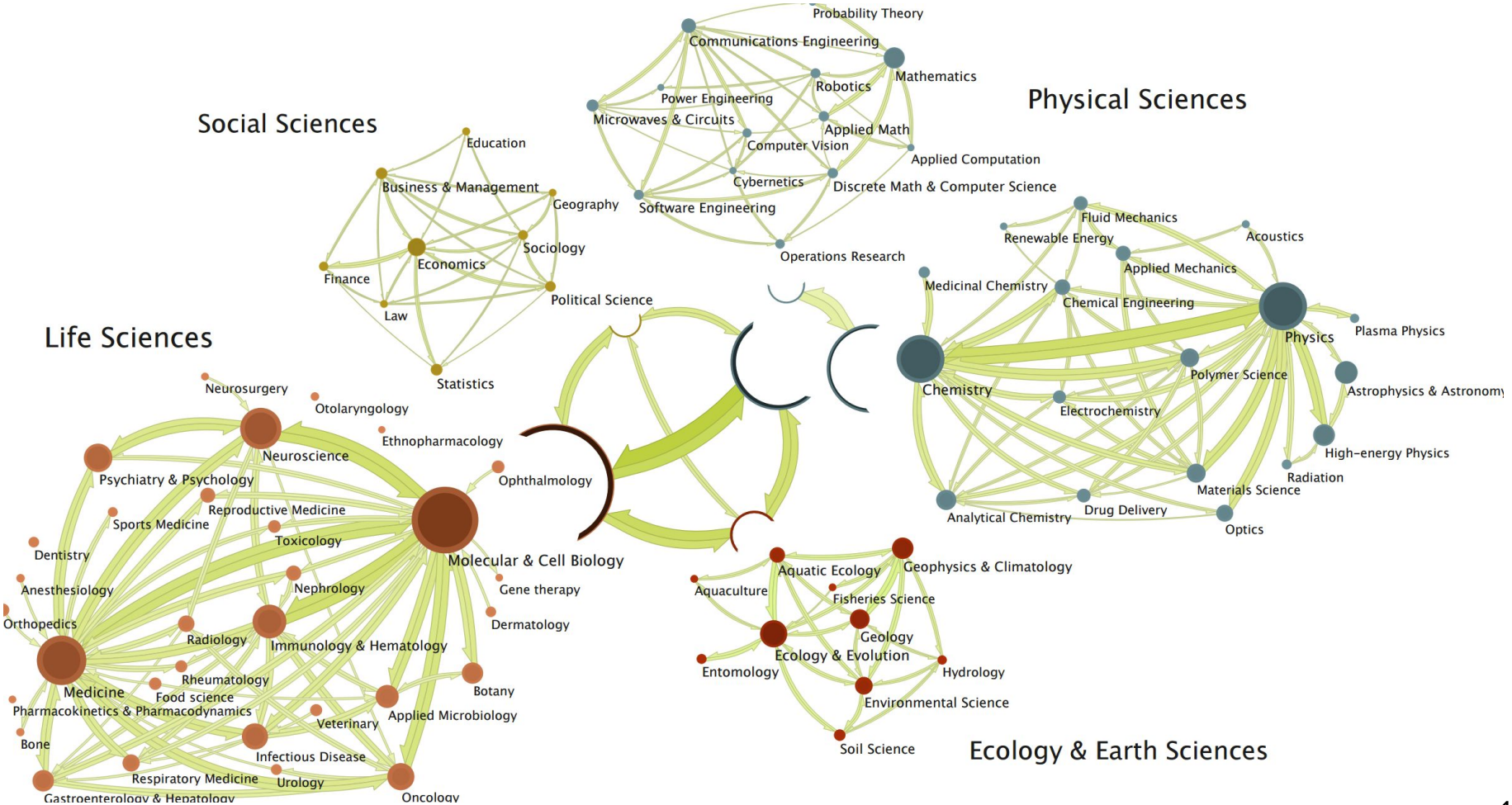
Science sampler

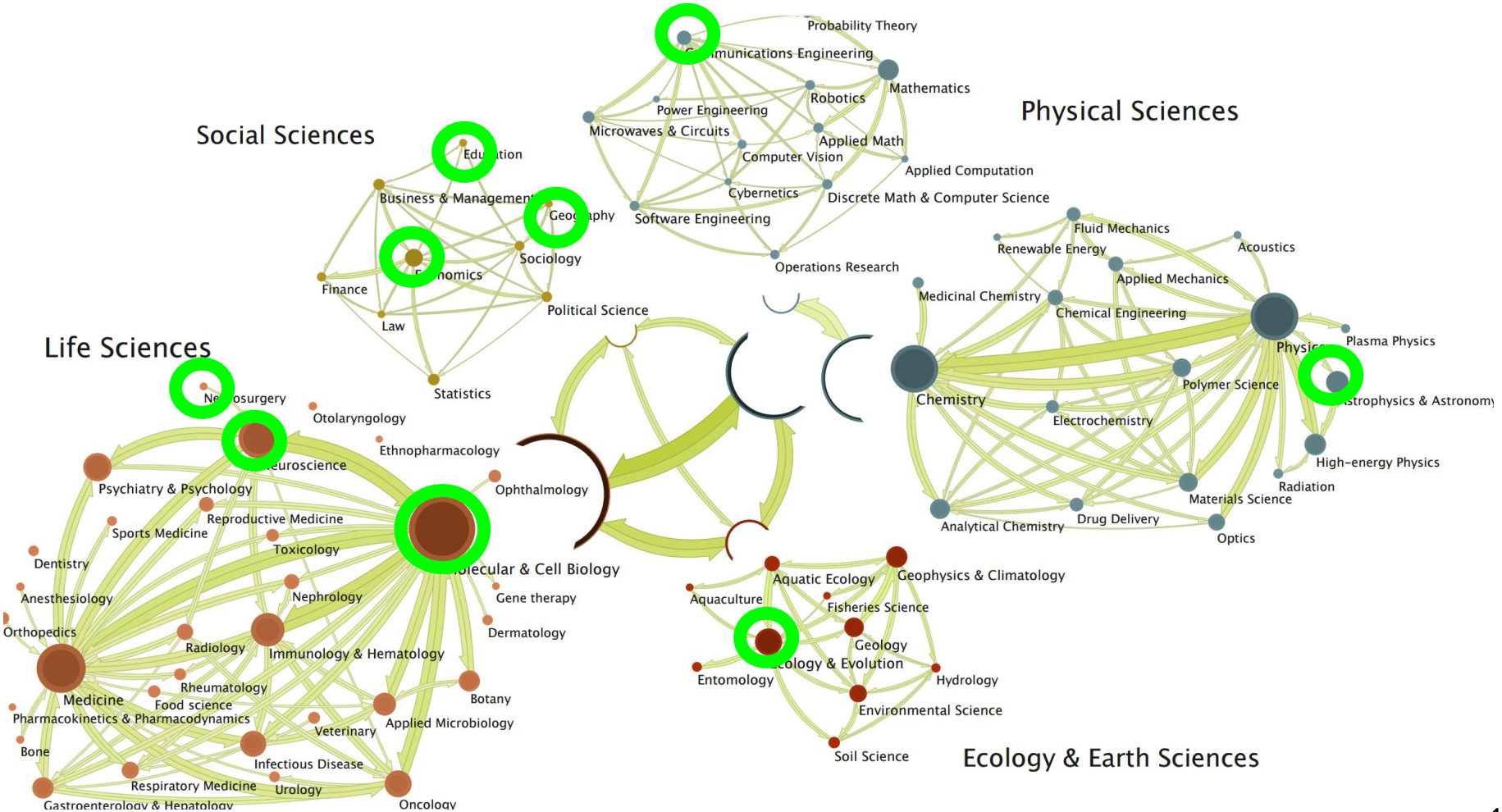
With apologies for the many projects we've not included. Selection here was almost random, weighted heavily by materials supplied by speakers here.

Science of Science






Social Sciences
Physical Sciences
Life Sciences
Ecology & Earth Sciences
Probability Theory
Communications Engineering
Mathematics
Robotics
Power Engineering
Microwaves & Circuits
Applied Math
Computer Vision
Applied Computation
Cybernetics
Discrete Math & Computer Science
Software Engineering
Operations Research
Education
Business & Management
Geography
Sociology
Economics
Finance
Political Science
Law
Statistics
Fluid Mechanics
Renewable Energy
Acoustics
Applied Mechanics
Medicinal Chemistry
Chemical Engineering
Physics
Plasma Physics
Polymer Science
Chemistry
Astrophysics & Astronomy
Electrochemistry
High-energy Physics
Radiation
Materials Science
Analytical Chemistry
Drug Delivery
Optics
Neurosurgery
Otolaryngology
Ethnopharmacology
Neuroscience
Psychiatry & Psychology
Ophthalmology
Reproductive Medicine
Sports Medicine
Toxicology
Dentistry
Molecular & Cell Biology
Anesthesiology
Nephrology
Gene therapy
Orthopedics
Dermatology
Radiology
Immunology & Hematology
Rheumatology
Botany
Medicine
Food science
Pharmacokinetics & Pharmacodynamics
Applied Microbiology
Veterinary
Bone
Infectious Disease
Respiratory Medicine
Urology
Gastroenterology & Hepatology
Oncology
Aquatic Ecology
Geophysics & Climatology
Aquaculture
Fisheries Science
Geology
Ecology & Evolution
Entomology
Hydrology
Environmental Science
Soil Science


Social Sciences
Physical Sciences
Life Sciences
Ecology & Earth Sciences
Probability Theory
Mathematics
Robotics
Power Engineering
Microwaves & Circuits
Applied Math
Computer Vision
Applied Computation
Cybernetics
Discrete Math & Computer Science
Software Engineering
Operations Research
Sociology
Finance
Political Science
Law
Statistics
Fluid Mechanics
Renewable Energy
Acoustics
Applied Mechanics
Medicinal Chemistry
Chemical Engineering
Plasma Physics
Polymer Science
Chemistry
Electrochemistry
High-energy Physics
Radiation
Materials Science
Analytical Chemistry
Drug Delivery
Optics
Otolaryngology
Ethnopharmacology
Psychiatry & Psychology
Ophthalmology
Reproductive Medicine
Sports Medicine
Toxicology
Dentistry
Anesthesiology
Nephrology
Gene therapy
Orthopedics
Radiology
Immunology & Hematology
Dermatology
Rheumatology
Medicine
Food science
Botany
Pharmacokinetics & Pharmacodynamics
Applied Microbiology
Veterinary
Bone
Infectious Disease
Respiratory Medicine
Urology
Gastroenterology & Hepatology
Oncology
Aquatic Ecology
Geophysics & Climatology
Aquaculture
Fisheries Science
Geology
Entomology
Hydrology
Environmental Science
Soil Science

Large Scale Genomics

- FASTQ files are mapped to a reference genome and converted to a BAM alignment file.
- BAM files can be mined for gene expression vectors that can be bundled into a gene expression matrix (GEM).
- GEMs are a stable data structure that can be mined for differentially expressed genes (DEGs) or used to construct Gene Co-expression Networks (GCNs)

Raw DNA Sequence →
Gene Interaction Graph

```
@SRR001666.1 071112_SLXA-EAS1_s_7:5:1:817:345 length=36
GGGTGATGGCCGCTGCCGATGGCGTCAAATCCCACC
+SRR001666.1 071112_SLXA-EAS1_s_7:5:1:817:345 length=36
IIIIIIIIIIIIIIIIIIIIIIIIIIIIII9IG9IC
```

OSG-GEM

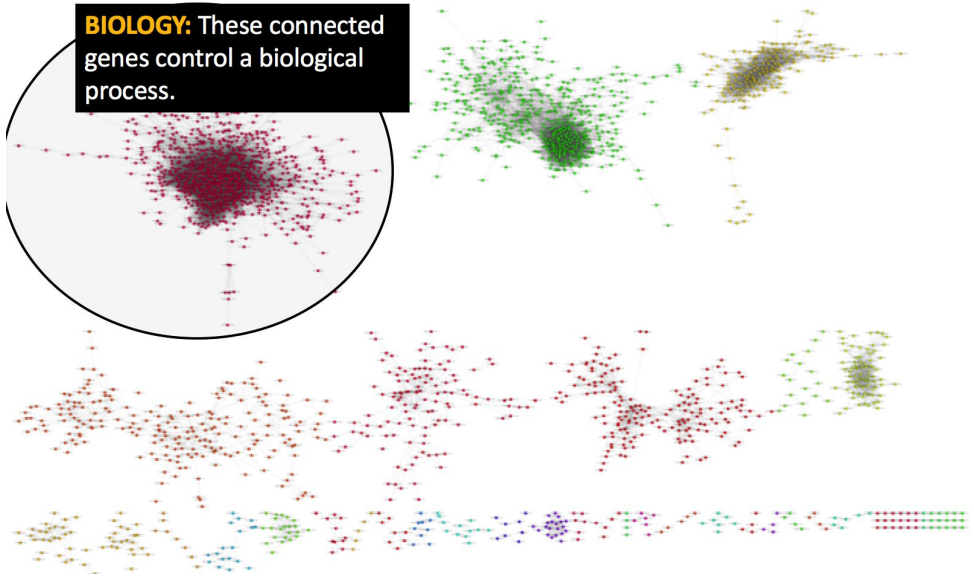
GeneID	ERR712353	ERR712355	ERR712362	ERR712366
GRMZM2G059856_T01.v6a	9.40	20.25	4.65	0.00
GRMZM5G811273_T02.v6a	0.00	0.38	0.00	0.45
GRMZM5G811273_T01.v6a	5.84	0.00	0.00	9.06
GRMZM2G059778_T01.v6a	0.45	1.65	0.00	0.88
GRMZM2G394757_T01.v6a	1.31	1.38	1.57	0.00
GRMZM2G093399_T01.v6a	9.00	1.08	1.00	0.65
GRMZM5G809743_T01.v6a	5.21	1.66	21.98	1.19
GRMZM5G833153_T01.v6a	2.98	30.50	19.33	15.75
AC177838.2_FGT012.v6a	0.29	0.24	0.00	0.00

OSG-KINC

Large Scale Genomics..

Genomics

COMPLEX GENETIC SYSTEMS WOVEN INTO GRAPHS



Gene Expression Matrix HPC Challenges

- Memory Intensive
  - Some Steps Require >16 GB RAM/dataset
- Storage Intensive
  - ~37 GB of intermediate files/dataset
- Dataset Size Varies Widely
  - 10M to 20M sequences/dataset
- Data Transfer Issues (Tera-/Petabytes)

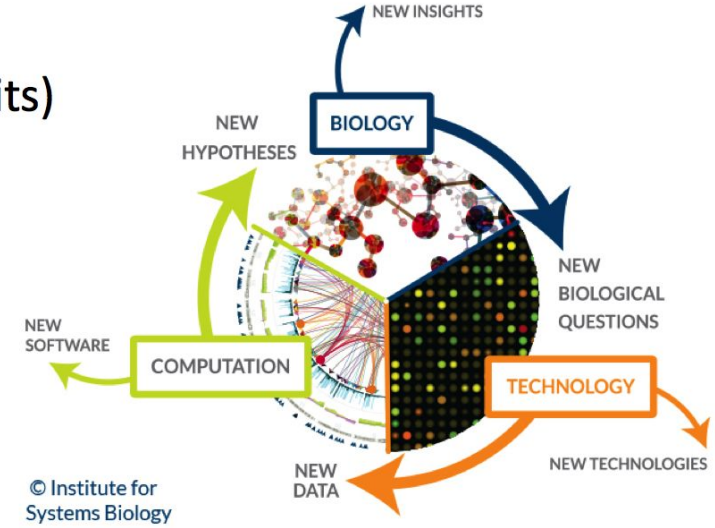
```
@SRR001666.1 071112_SLXA-EAS1_s_7:5:1:817:345 length=36
GGGTGATGGCCGCTGCCGATGGCGTCAAATCCCACC
+SRR001666.1 071112_SLXA-EAS1_s_7:5:1:817:345 length=36
IIIIIIIIIIIIIIIIIIIIIIIIIIIIII9IG9IC
```

William Poehlman, Alex Feltus • Clemson University •• Stephen Ficklin, Washington State University

Student Training w/ OSG

Genomics

- Introduction to Systems Biology
- WSU HORT 503 Special Topics (3 credits)
- Offered Fall 2017
- 9 Students
  - Crop & Soil Sciences
  - Molecular Plant Sciences
  - Entomology
  - Biological Systems Engineering
- Why OSG?
  - Expose students to national computing resources
  - Anticipate large-scale genomics of the future.

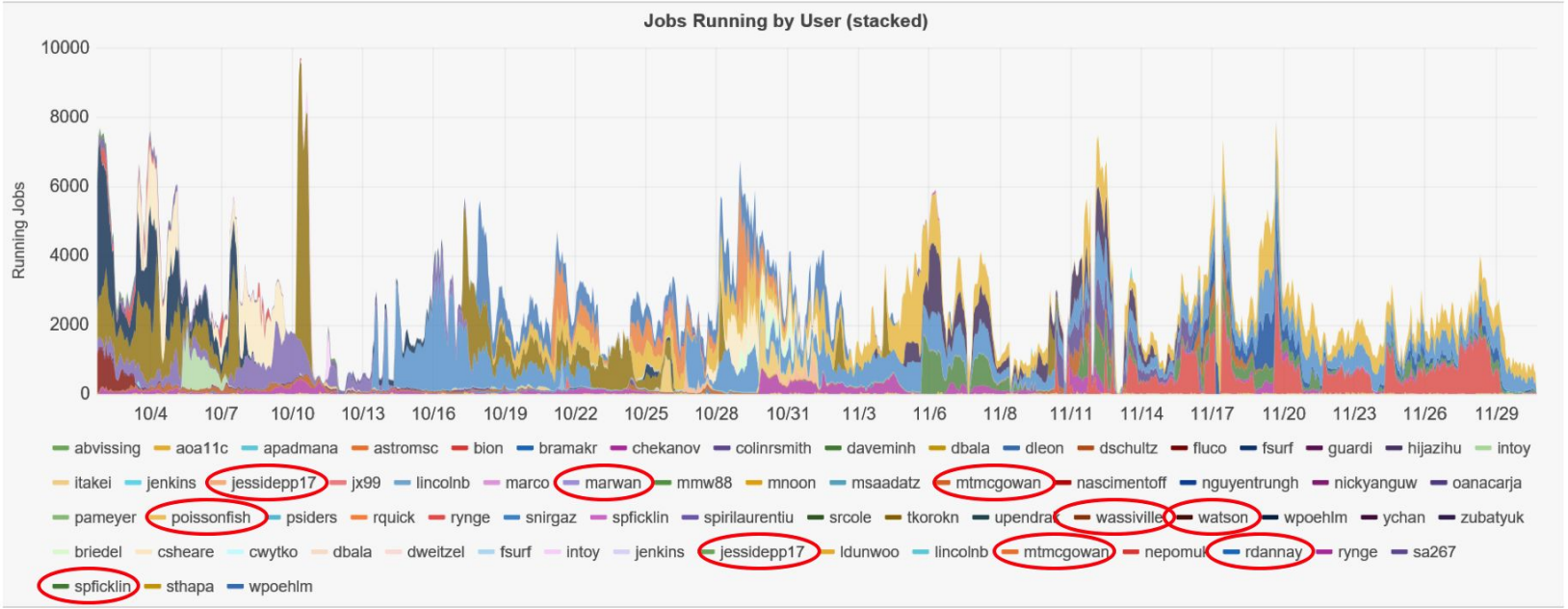


WASHINGTON STATE UNIVERSITY

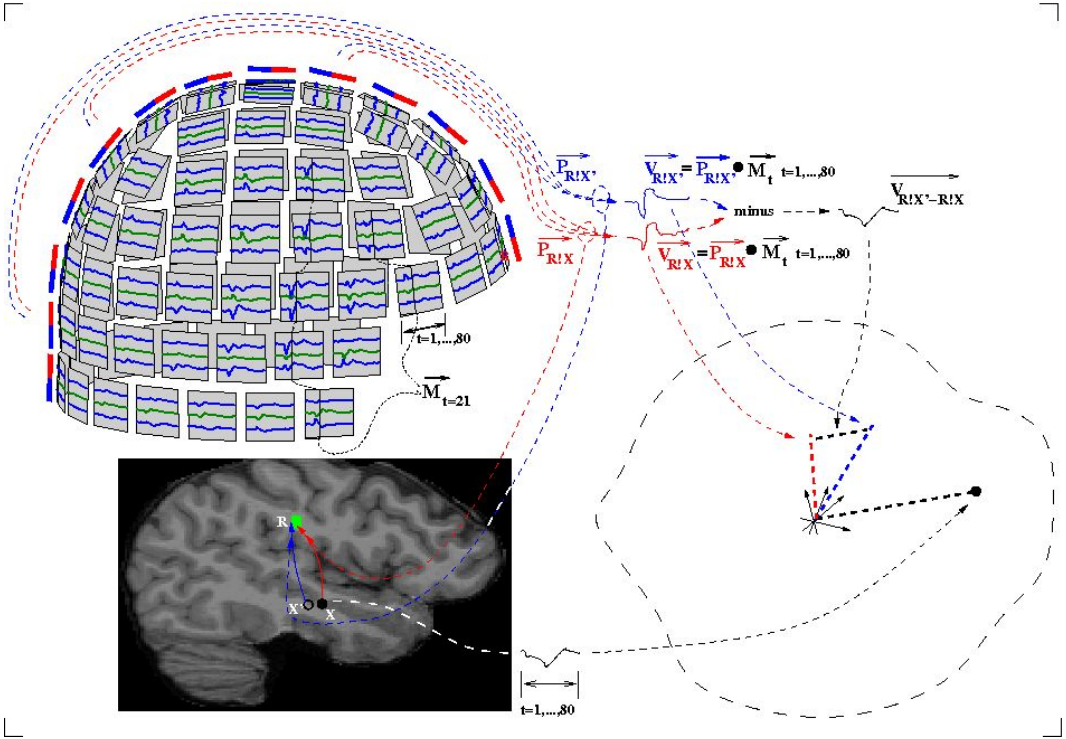
Student Training w/ OSG

Genomics

- Success: SysBioEdu Usage from Oct-Dec 2016
  - Students launched jobs and they ran!



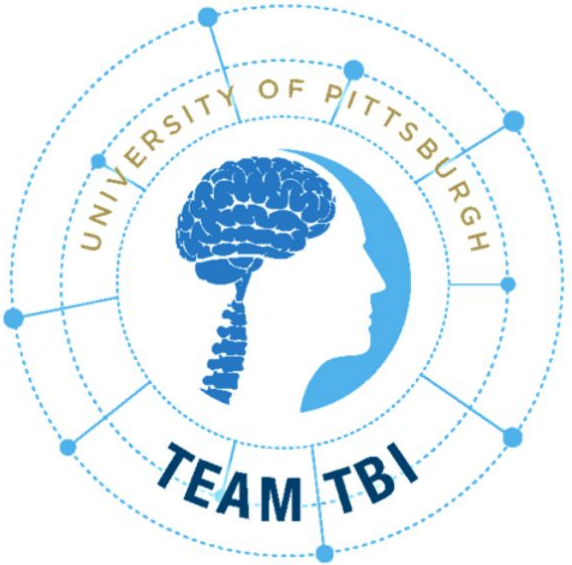
Functional Neuroimaging



Functional Neuroimaging..

- Don Krieger has been working with TEAM TBI at the University of Pittsburgh
  - **T**argeted **E**valuation, **A**ction and **M**onitoring of **T**ramatic **B**rain **I**njury

- TEAM TBI investigates the complexity of brain injury, and how targeted interventional strategies may improve outcome and function.



TeamTBI Functional Neuroimaging

David Okonkwo, Jim Becker, Sue Beers, Mickey Collins, Anthony Kontos, Malcolm McNeil, Walt Schneider

TeamTBI Functional Neuroimaging

David Okonkwo, Jim Becker, Sue Beers, Mickey Collins, Anthony Kontos, Malcolm McNeil, Walt Schneider

Supported by: Department of Defense, Open Science Grid (NSF,DOE), Extreme Computing Environment for Science and Engineering (NSF).

TeamTBI Functional Neuroimaging

We record the magnetoencephalogram (MEG) from an array of 306 magnetometers sampled at 1 KHz typically for 30 minutes.

TeamTBI Functional Neuroimaging

We record the magnetoencephalogram (MEG) from an array of 306 magnetometers sampled at 1 KHz typically for 30 minutes.

We utilize a unique deconvolution solver to identify and validate ($p<10^{-12}$) each neuroelectric source within the brain with mm resolution.

TeamTBI Functional Neuroimaging

We record the magnetoencephalogram (MEG) from an array of 306 magnetometers sampled at 1 KHz typically for 30 minutes.

We utilize a unique deconvolution solver to identify and validate ($p<10^{-12}$) each neuroelectric source within the brain with mm resolution.

The solver requires about 40 core hours per second of data. It typically yields about 500 simultaneously active sources for each 40 msec step through the data stream, i.e. 20,000,000+ sources from a single 30 minute recording session.

TeamTBI Functional Neuroimaging

We record the magnetoencephalogram (MEG) from an array of 306 magnetometers sampled at 1 KHz typically for 30 minutes.

We utilize a unique deconvolution solver to identify and validate ($p<10^{-12}$) each neuroelectric source within the brain with mm resolution.

The quantity and resolution of these measurements is unprecedented.

TeamTBI Functional Neuroimaging

In December, we gained access to the CamCAN lifespan normative cohort in collaboration with the University of Cambridge (UK). These recordings were obtained from 628 volunteers ages 18-87.

TeamTBI Functional Neuroimaging

In December, we gained access to the CamCAN lifespan normative cohort in collaboration with the University of Cambridge (UK). These recordings were obtained from 628 volunteers ages 18-87.

Processing has been completed on 100 resting records using OSG including reserved glideins running on SDSC's Comet and PSC's Bridges.

TeamTBI Functional Neuroimaging

Consistent patterns of activation across 155 brain regions were identified from the CamCAN recordings with eigenvector analysis.

TeamTBI Functional Neuroimaging

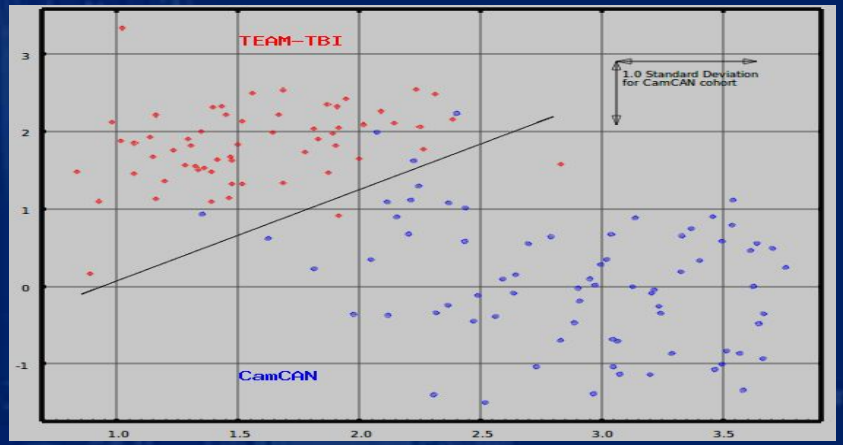
Consistent patterns of activation across 155 brain regions were identified from the CamCAN recordings with eigenvector analysis.

The brain patterns found in MEG recordings from volunteers with history of head injury are sufficiently different to classify them with 95% accuracy.

TeamTBI Functional Neuroimaging

Consistent patterns of activation across 155 brain regions were identified from the CamCAN recordings with eigenvector analysis.

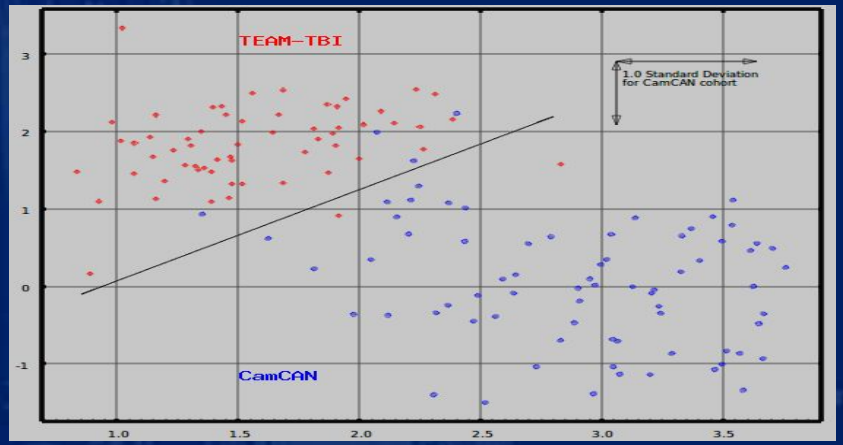
The brain patterns found in MEG recordings from volunteers with history of head injury are sufficiently different to classify them with 95% accuracy.



TeamTBI Functional Neuroimaging

Consistent patterns of activation across 155 brain regions were identified from the CamCAN recordings with eigenvector analysis.

The brain patterns found in MEG recordings from volunteers with history of head injury are sufficiently different to classify them with 95% accuracy.

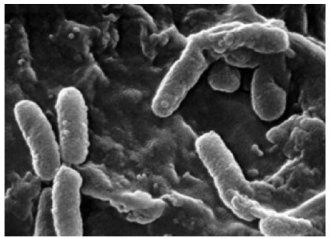


One of these two brain patterns used to classify the cohorts is specific to depression, the other to somatization. These are common symptoms of PTSD.

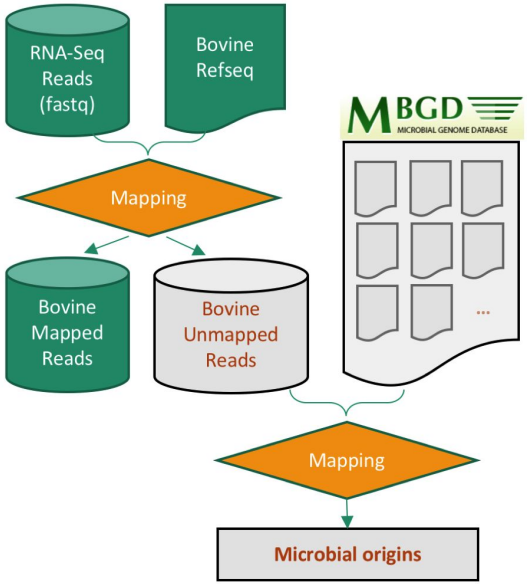
Large Scale Metagenomics.

Motivation

- In a pervious project, we have isolated exosomes from the commercial cow's milk and assessed the bovine molecules inside the exosomes.
- Moreover, we also found many unmapped reads are from microbial species.
- Thus, we designed a follow-up study to **understand the origin of microbial sequences in bovine milk exosomes.**
  - **Metagenomics analysis**



Pseudomonas cedrina



Large Scale Metagenomics..

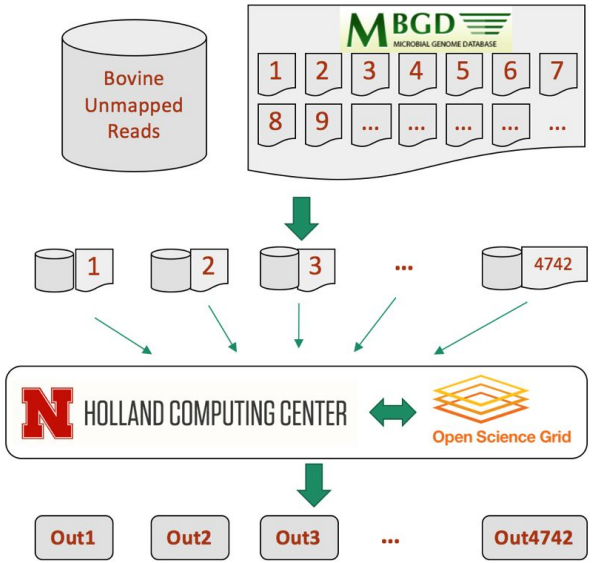
Computational challenges

- A large number of target genomes
  - 4,742 genomes (size: 100KB ~ 58MB)
- Six samples contain over 100 million of bovine unmapped reads
  - Increasing the computing time
- In total, 6 x 4,742 = ***28,452*** mapping tasks

Question: Where to execute this many of jobs?
- Impossible for the lab-server (32 cores)
- Long pending time if submitted it to HCC clusters
  - Dynamic priority scheduling of users/groups
  - More jobs completed -> longer queue time

Perfect Fit of Open Science Grid (OSG)

- The tasks are independent to each other
- Limited file transfer
  - Total size of transferred files ~1GB
- Small memory consumptions
  - Memory < 2GB
- Short running time for each task
  - Maximum: 3 hours (HCC@UNL-Crane)
- Software is available on OSG
  - Pre-installed Bowtie and Tophat
  - No further configuration needed



Large Scale Metagenomics…

Results

- Several microbial species were identified in bovine milk exosomes:
  - *Pseudomonas fluorescens, Pseudomonas chlororaphis, Pseudomonas poae, Enterobacter cloacae,* etc.
- Although some species have been reported in cow's milk before, this is the first time of identifying microbial sequences in milk exosomes
  - Potential transportability to other species through exosomes
- Based on the findings from this analysis, we have designed two experiments to further our understanding in this subject

Counterfactual Analysis.

- Economic analysis & public policy
- Considering "what if" scenarios in microeconomics
- Simulate firm/consumer behaviors

Using OSG to Evaluate Policy

We need to solve

$$\frac{\partial \Pi_{jt}}{\partial p_{jt}} + \beta \left[\frac{\partial \boldsymbol{s}_t}{\partial p_{jt}} \right]' \mathbb{E}_t \left[\frac{\partial V_j(\boldsymbol{s}_t, \boldsymbol{X}_t)}{\partial \boldsymbol{s}_t} \right] = 0,$$

for every firm and combination of $\boldsymbol{s}_t$ and $\boldsymbol{X}_t$.

Need to compute $\mathbb{E}_t \left[\frac{\partial V_j(\boldsymbol{s}_t, \boldsymbol{X}_t)}{\partial \boldsymbol{s}_t} \right]$. How? Forward simulation.

OSG is what makes this possible for large state spaces.

Counterfactual Analysis..

Using OSG to Evaluate Policy

Outcome: making people more active decreases prices and it does not affect returns.

Case	Mean and 95% CI
Base simulation	6.195% [6.181%,6.210%]
No enrollment cost	3.666% [3.660%,3.671%]
No decision cost	3.837% [3.833%,3.842%]
No switching costs	2.607%

Note: The table reports the mean expected fees and 95% confidence intervals for the different scenarios under study using $10,000$ random initial states.

Conclusions

1. Policy/transaction evaluation is critical in modern microeconomics.
2. Evaluations often require simulating consumer/firm behavior over a large state space.
3. In particular in the case of dynamic games, OSG could become a crucial tool, making the difference between being able to do something and not.

Simulating Source Coding.

- Data deluge - much of it mobile traffic
- Optical data compression
- Important for digital space and satellite communication & wireless data transmission

Monte Carlo Simulation for Next Generation Source and Channel Coding on OSG Connect

Ahmad Golmohammadi*, David G. M. Mitchell*,
Joerg Kliewer †, and Daniel J. Costello, Jr. ‡

* Klipsch School of ECE, New Mexico State University

† Dept. of Electrical and Computer Engineering, New Jersey Institute of Technology

‡ Dept. of Electrical Engineering, University of Notre Dame

March 7, 2017

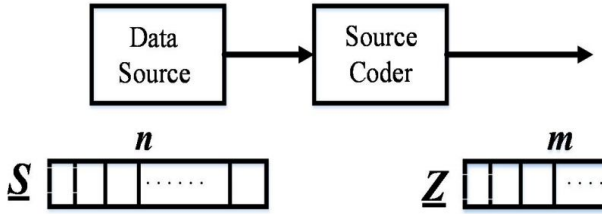
1

Simulating Source Coding..

- Whole system simulations: transmitter, decoder, receiver & stochastic noise
- Data compression & reconstruction algorithms

Data Compression

❑ In order to handle the vast amounts of data society will produce, we need efficient and low-complexity algorithms to reliably compress and reconstruct data.



Compression rate

$$R = \frac{m}{n}$$

❑ We want to reconstruct the sequence with as little distortion as possible with a practical scheme (lossy source coding)

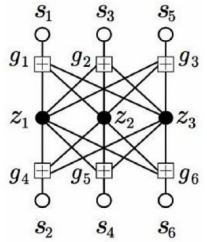
❑ The ultimate limit of compression is known (Shannon 1958)

Simulating Source Coding...

- Sparse graphs can approach fundamental limits
- To verify the results, large Monte Carlo samples needed - "not possible without the OSG"

LDGM Codes for Lossy Source Coding

- We construct codes for lossy source coding using a small structured graph (protograph)
- A large graph can be obtained from a protograph by graph lifting with lifting factor M
  - The graph is copied M times and the edges randomly permuted following the graph structure
- Low complexity algorithms based on belief propagation can be defined for the sparse graph
  - Here, messages are passed iteratively forward and backward in the graph until we converge to a codeword $\underline{\boldsymbol{z}}$



(3, 6)- regular

Evolving Strategies for Life.

- Understanding evolution at molecular scale in DNA with combination of mathematical modeling and simulation
- How quickly does a genome fix a mutation?
- Role of randomness versus natural selection?

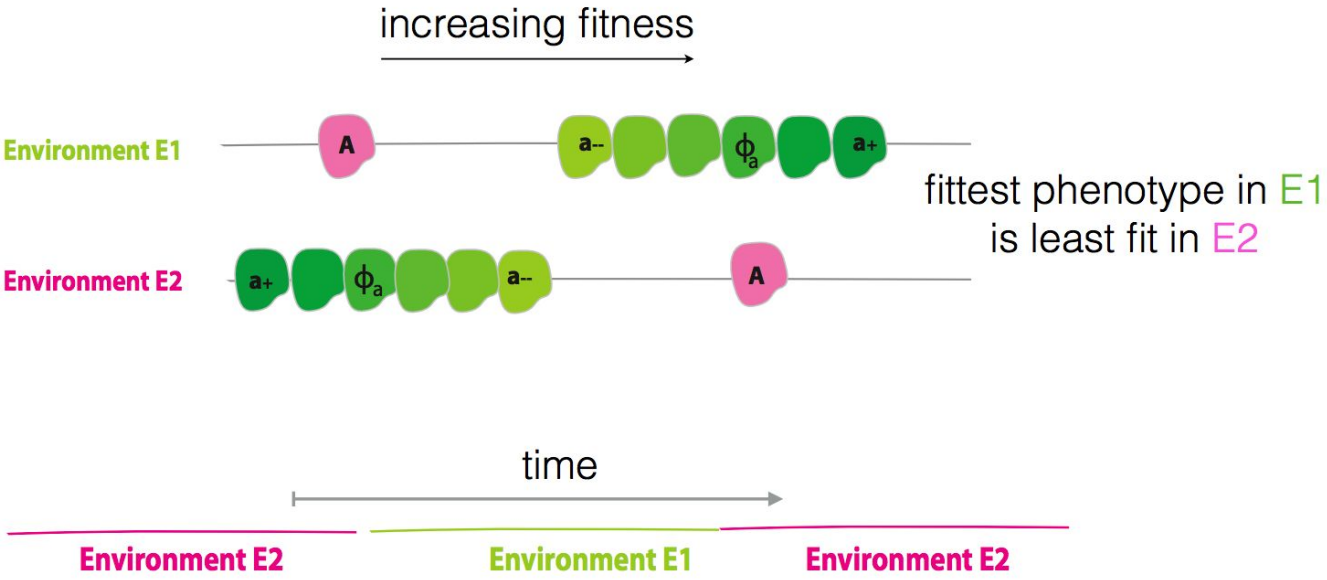
Environmental variation is commonplace yet unpredictable across biological systems from the adaptive immune system, the microenvironment in cancerous neoplasms, to populations of pathogens under drug pressure.

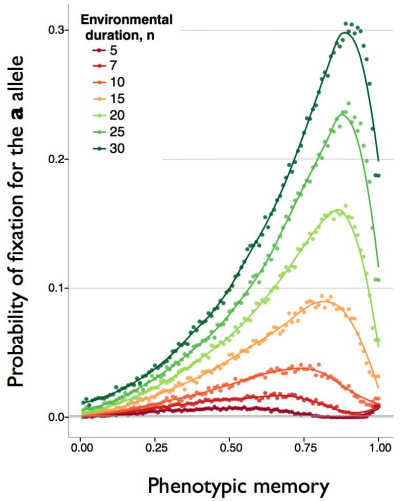
How do populations survive environmental stochasticity? How do they manage to persist and keep one's footing on an ever-changing landscape?

Can organisms prepare for this environmental stochasticity?

Can evolution prepare populations for this environmental stochasticity?

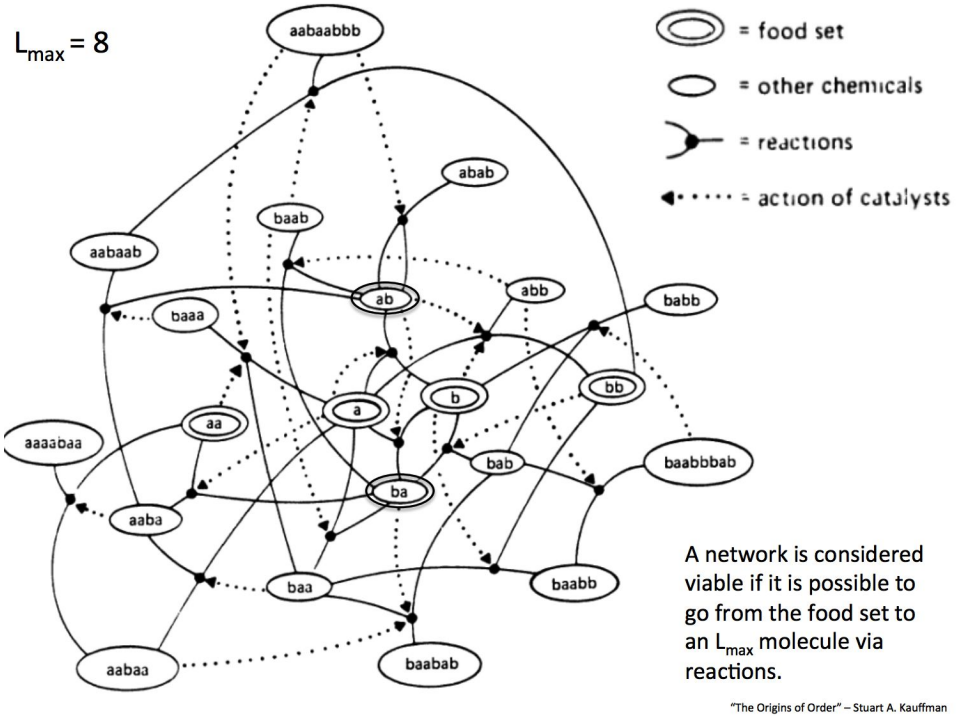
Evolving Strategies for Life..





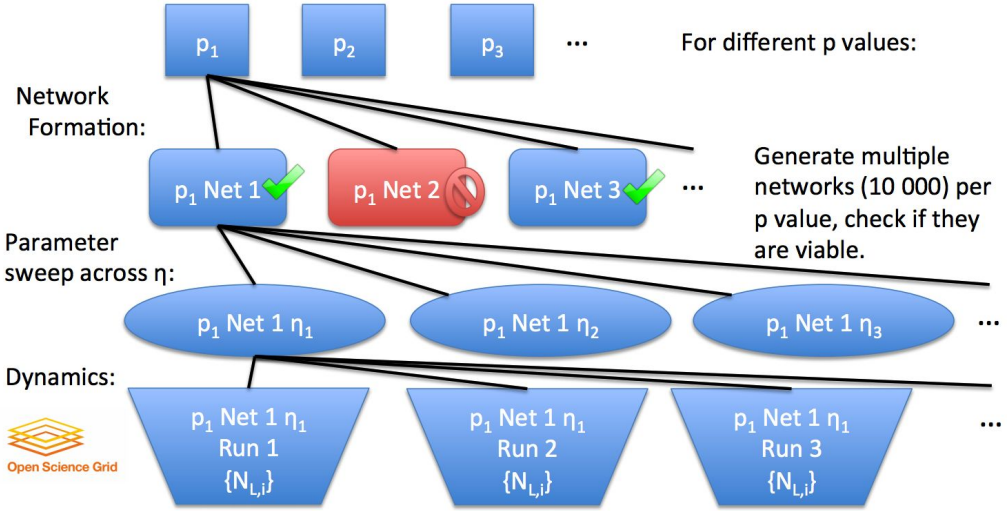
Models of Prebiotic Evolution

- Protein first origin of life model
- Network of interacting molecules assumed to be polymers
- Perhaps solve Eigen's paradox (low probability of randomly constructing "starter gene")



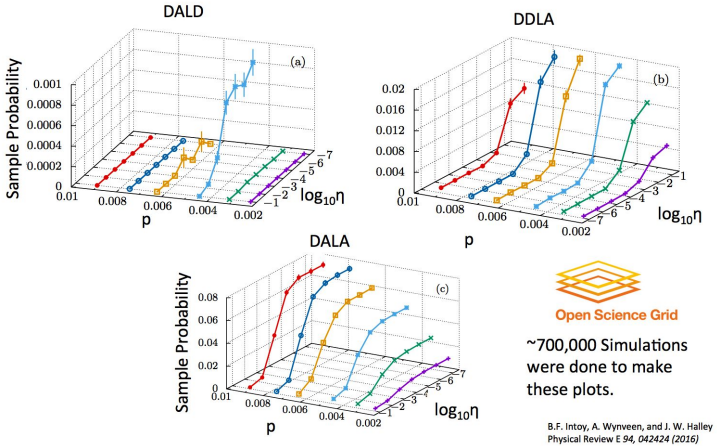
Models of Prebiotic Evolution..

Simulation General Structure



- Do multiple dynamic simulations with random initial conditions using a given viable network generated by parameter p combined with reaction rates and diffusive value η.
- A steady state is then reached with polymer length and spatial distribution $\{N_{L,i}\}$.
- Analyze the $\{N_{L,i}\}$'s to determine whether the run was lifelike or not.

Probabilities of DALD, DDLA, DALA states as a function of p and η



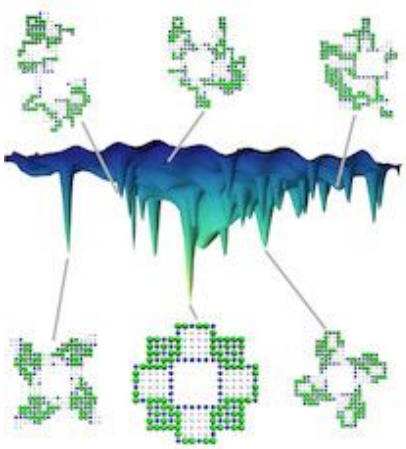
Open Science Grid

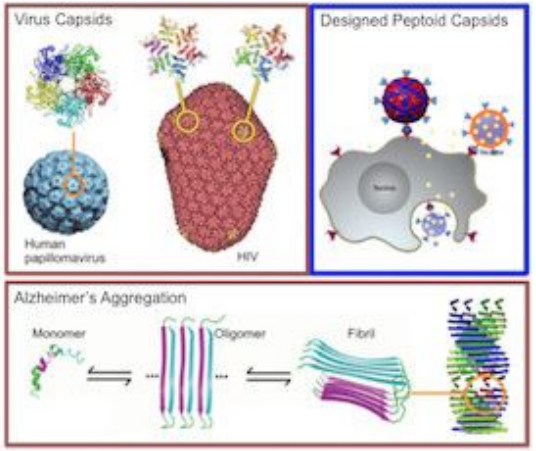
~700,000 Simulations were done to make these plots.

B.F. Intoy, A. Wynveen, and J. W. Halley Physical Review E *94, 042424 (2016)*

Protein Evolution

Understand the fundamental physical bottlenecks and dynamical behavior of protein evolution. Important questions include the extent of dominant pathways (convergent evolution) and phase transitions in evolutionary rates (punctuated equilibrium). These principals and their structural underpinnings can also be used to inform rational design of antibiotics that exploit bottlenecks in pathogen mutational response.



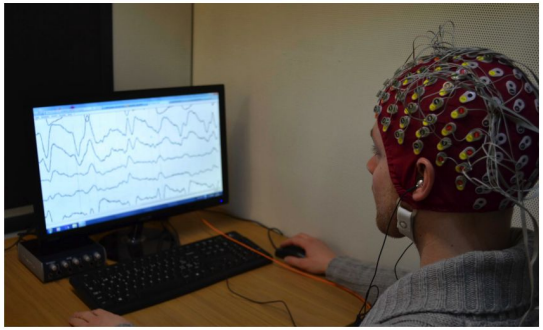


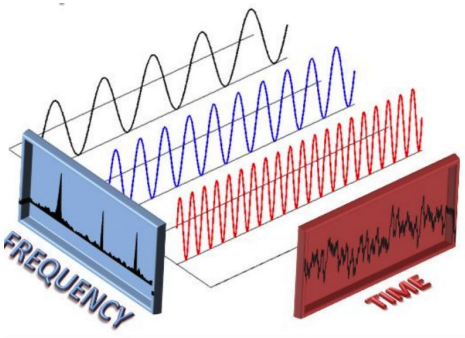
Analysis of Brain Rhythms

Neuroscience

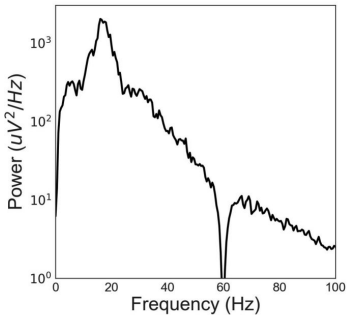
Large-scale brain recordings

- Healthy humans
- Patients
- Animals





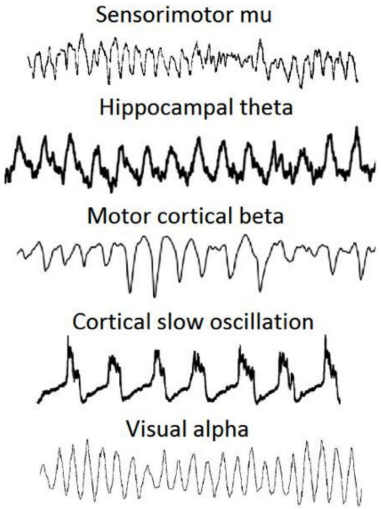
- Sampling: 500-30000 Hz
- Duration: 1 hour - 1 week
- Channels: 1 - 250+
- Several GB per subject

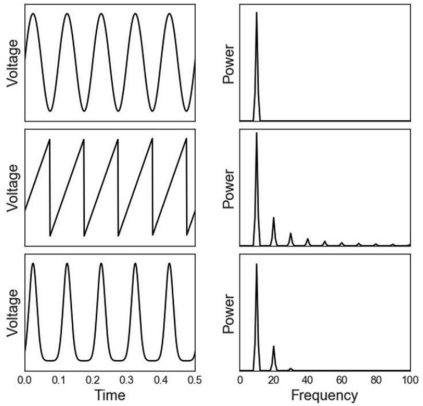


Analysis of Brain Rhythms..

Neuroscience

Fourier Transform-based analysis





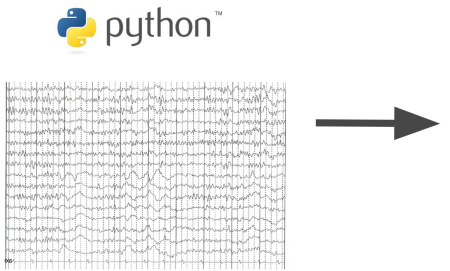
Free supercomputing for research: A tutorial on using Python on the Open Science Grid

Jan 3, 2017

Supercomputing resources typically cost money, but the Open Science Grid (OSG) provides high-throughput computing to any researcher in the US **for free**. Briefly, OSG users can run jobs on servers owned by dozens of academic institutions, whenever those servers are not actively running a job for its owners.

Besides cost, a second major barrier to entry for those who are new to supercomputing (or, specifically, Condor) is the necessary troubleshooting before we can actually run our jobs. The purpose of this tutorial is to provide a complete example for running Python jobs on the OSG. This example is nontrivial, in that it includes multiple data sets, public libraries (e.g. scipy), private libraries, and analyzing output. In complement to this tutorial, the OSG has tutorials, a structured class, and extremely helpful online support when you get stuck.

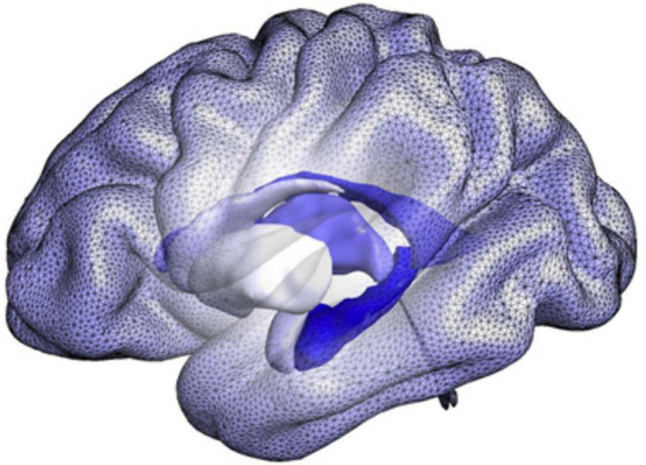
This tutorial goes through the steps of manually connecting to and running commands on the remote server, but see the **Fabfile** section at the bottom for how this can be automated on your local machine.



A FreeSurfer Workflow Service

Neuroscience

- Widely used software suite for analysis of human brain MRI scans.
- Neurophysiology of depression, examining possible anatomical differences involved in ADHD, and studying autism



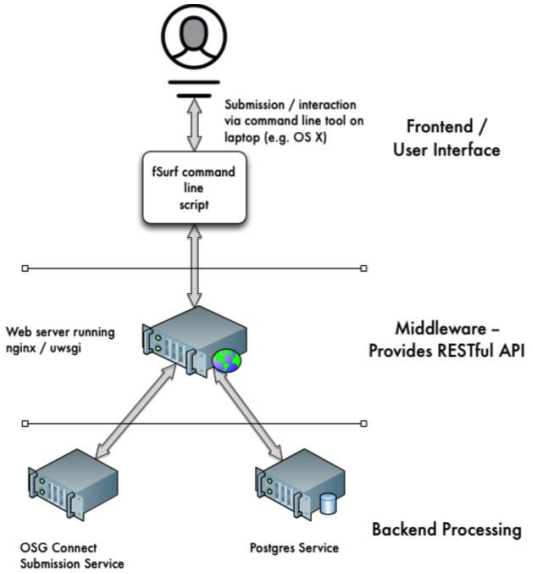
Computational model of the cortical and subcortical brain structures that form the basis of the BrainPrint, a system for representing the whole brain based on the shape, rather than the size of structures. (Martin Reuter, PhD, and Christian Wachinger, PhD, Martinos Center for Biomedical Imaging)

A FreeSurfer Workflow Service

Neuroscience

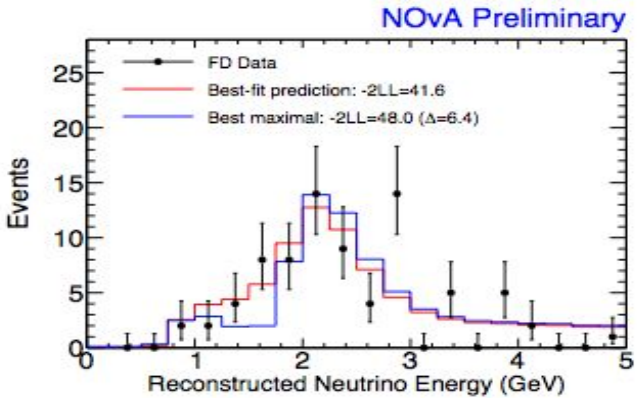
- Working with Don Krieger (Pittsburgh) to develop an OSG-based execution service
- Uses Pegasus
- Handles "standard" transforms and user options
- To be released this week!

- Front-end user interface
  - Command line script
    - Communicates to the execution service via RESTful-API commands
    - Allows user to submit, view, and remove workflows as well as download results
  - Middleware
    - NGinx / uwsgi server provides RESTful API
- Backend services (running on a single VM)
  - Postgres database
    - Tracks workflow status, locations of inputs and results
    - Stores user information
  - Pegasus submission service (using same infrastructure as OSG Connect) to



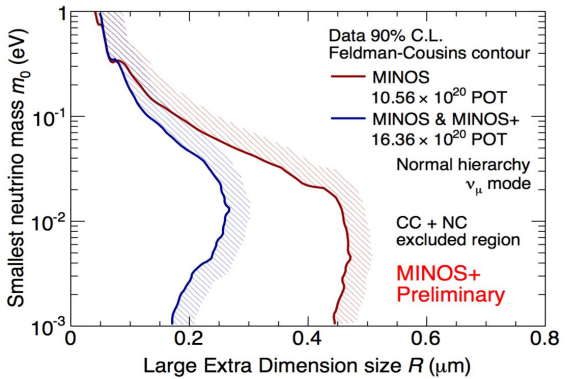
VOs

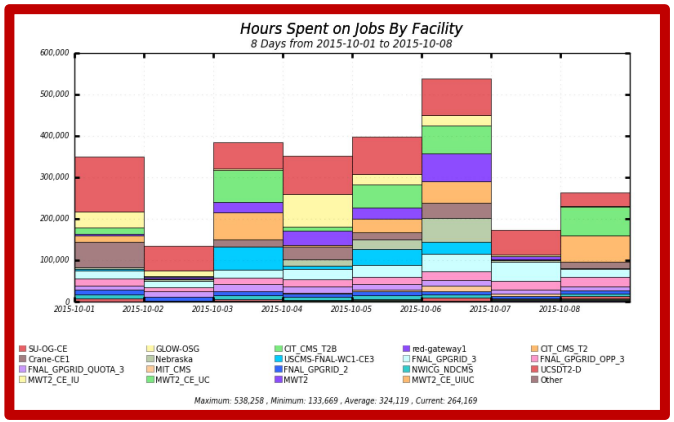
VO Highlights: From the smallest scales…



NOvA: Fermilab-based neutrino experiment

MINOS+: limits on LEDs



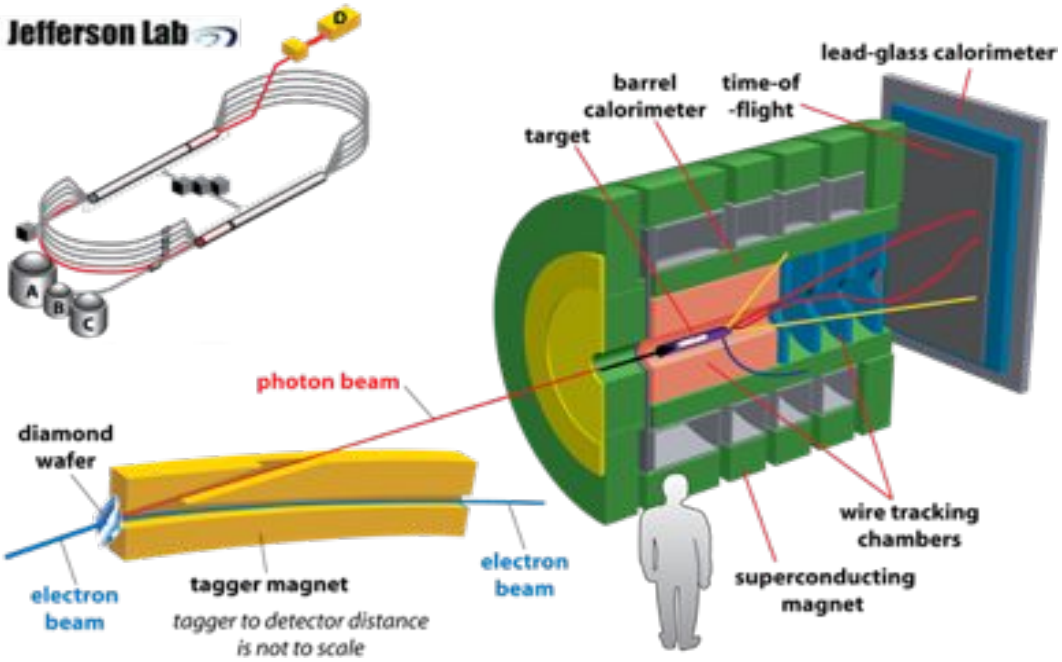


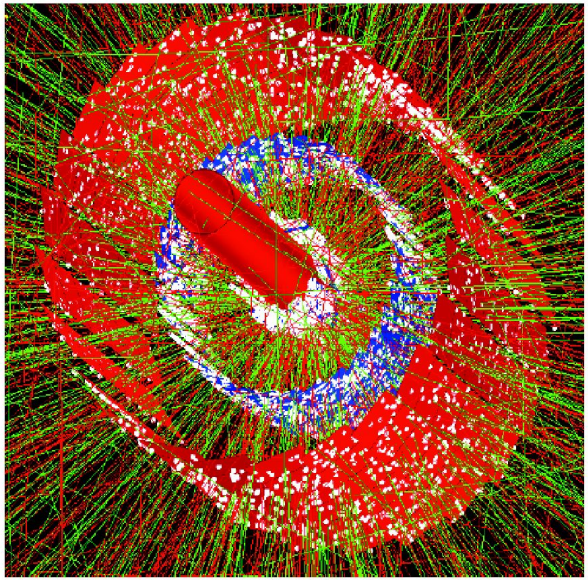
Mu2e: Lepton-flavor violation experiment

Nearly 60M opportunistic hours on OSG and counting

>500,000 in one day!

VO Highlights: From the smallest scales...



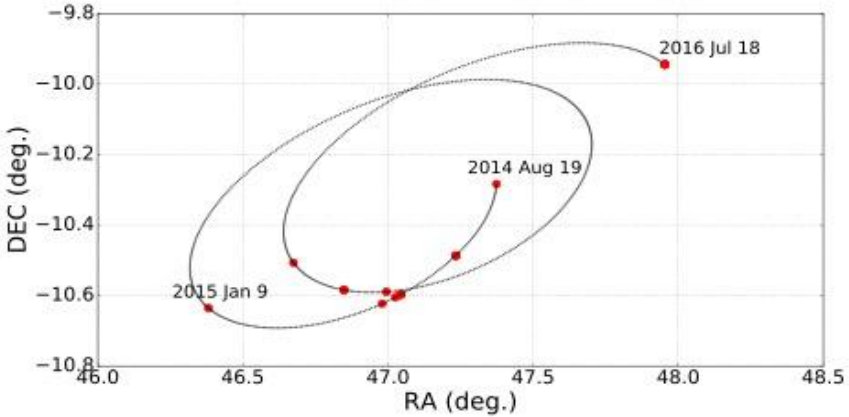


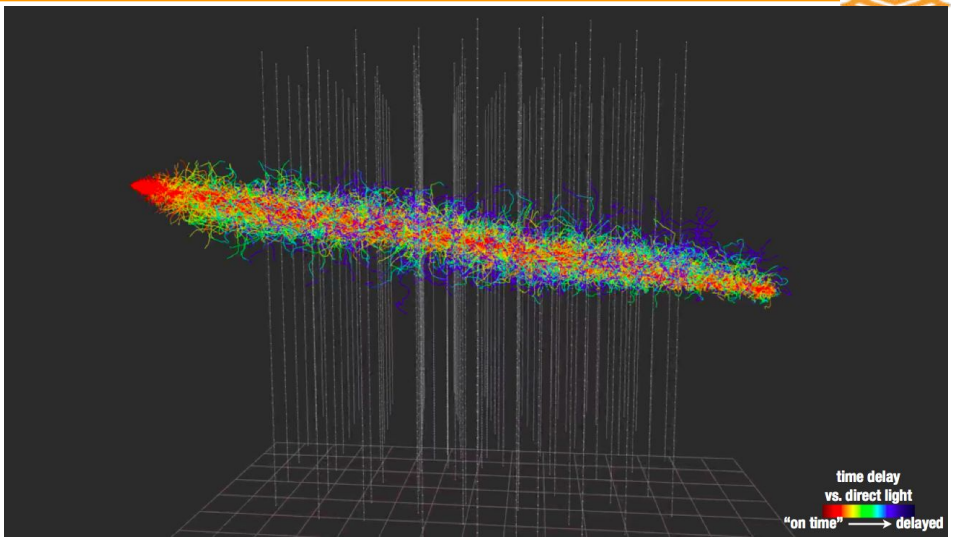
Au+Au event

GlueX: probing exotic mesons predicted by LQCD

STAR: Heavy Ion Physics

...to the largest...

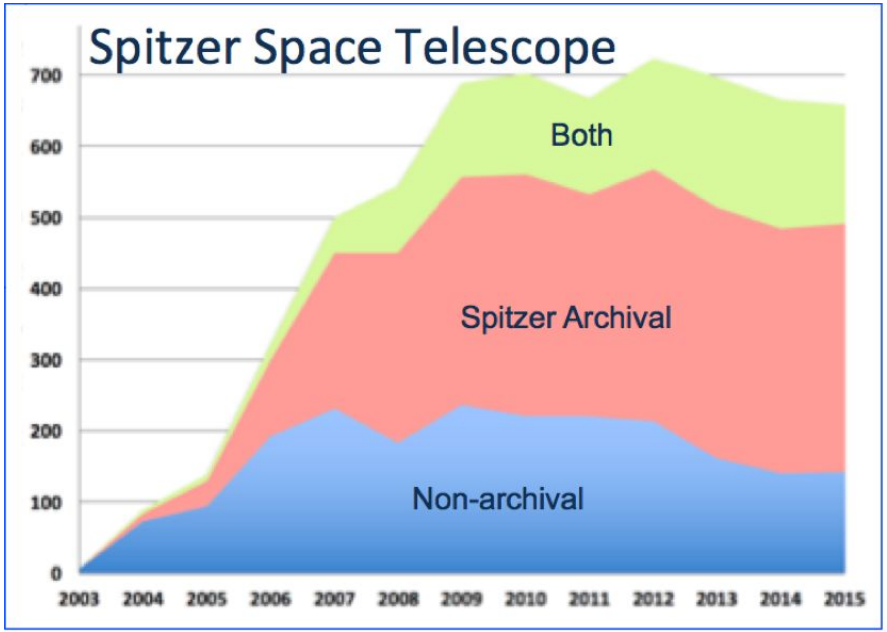




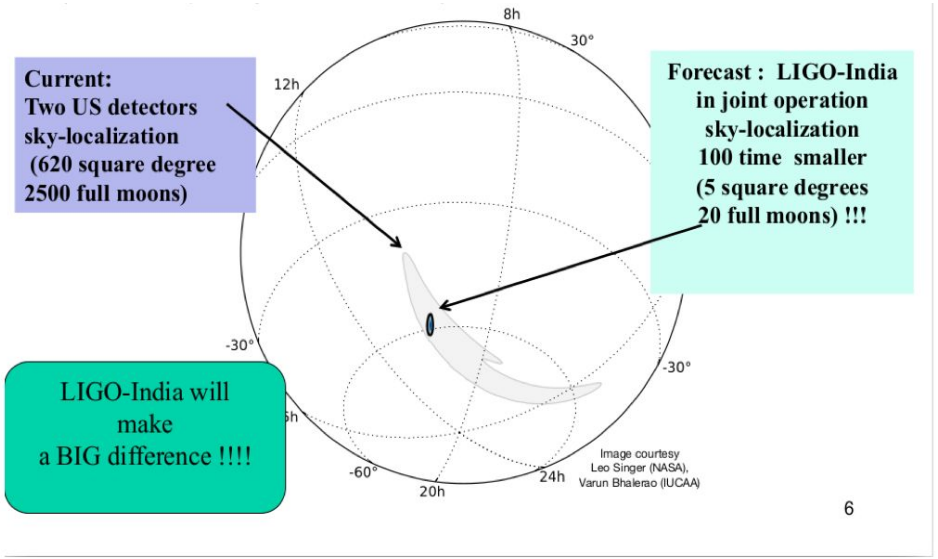
Dark Energy Survey:
Discovery of dwarf planet-
Second-most distant known object in solar system
Techniques applied to ongoing Planet 9 search

Ice Cube:
Neutrino Observatory, also sensitive to extremely high energy cosmic rays

…the completed to the still in planning…



Infrared Processing and Analysis Center:
NASA's archive for a host of IR/sub-mm astronomy missions, galaxy catalogs, Keck Observatory, and more!



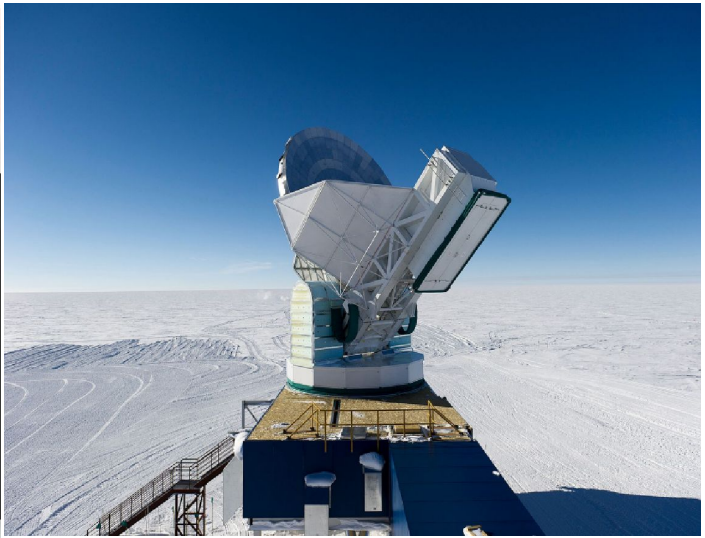
[Tarun Souradeep (2016), Beyond Gravitational wave discoveries with LIGO-India]

LIGO India:
Additional detector will greatly Improve localization of gravitational wave sources

...and working in all corners of the globe.

VERITAS:
4 12m Cerenkov telescopes for gamma ray astromony: Arizona, USA

XENON1T:
Dark matter detector at Gran Sasso National Laboratory, Italy



South Pole Telescope:
Microwave-millimeter telescope

And in space!

Alpha Magnetic Spectrometer (AMS) mounted at the ISS
Photo credit: NASA

OSG User Support

Team

http://support.opensciencegrid.org/

opensciencegrid help desk

Welcome
Login Sign up

Home | Solutions

How can we help you today?

Enter your search term here... SEARCH

New support ticket

Check ticket status

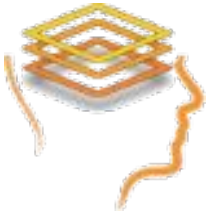
User Support

Training

Intro to HTC on OSG Connect in Synthesis Center tomorrow

software carpentry
osg connect

Thank you!



opensciencegrid

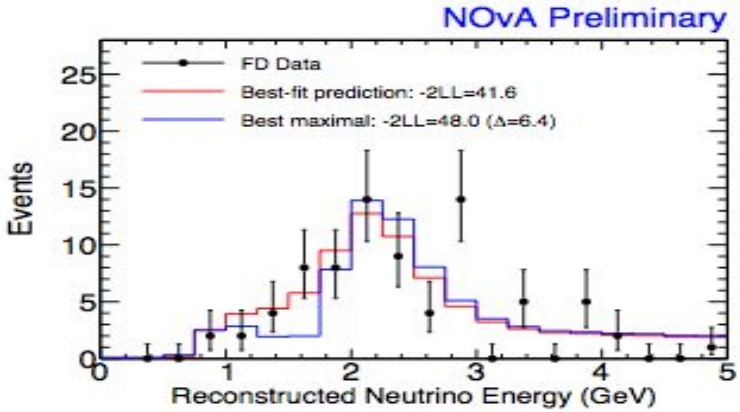
user-support@opensciencegrid.org

support.opensciencegrid.org

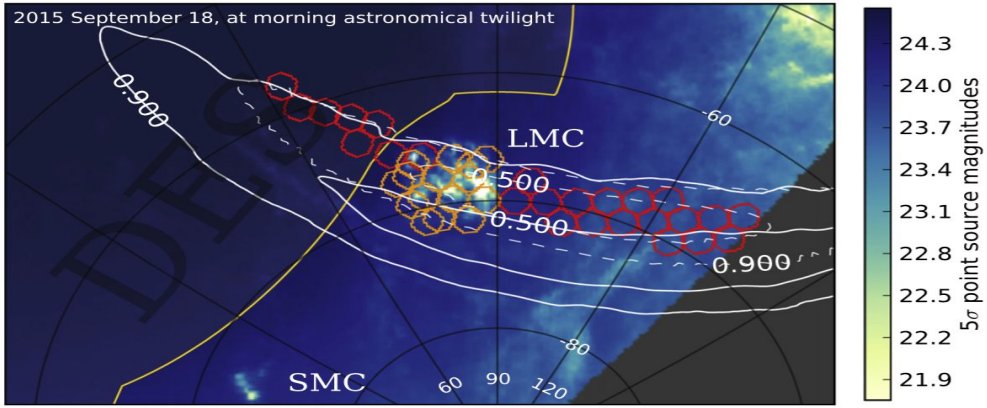
www.opensciencegrid.org/links

Extra slides

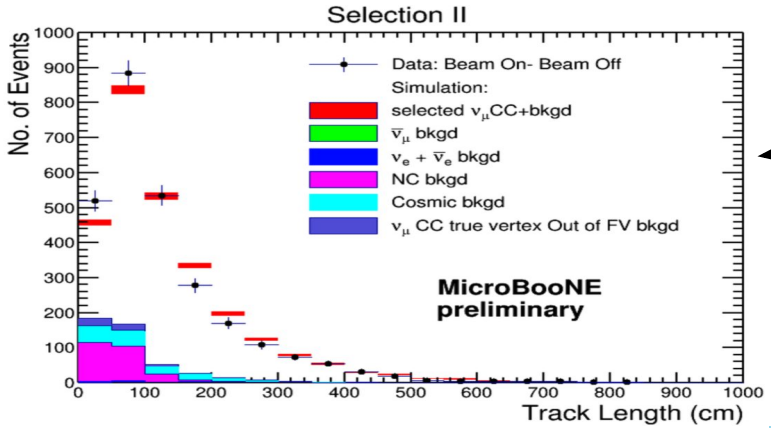
Selected results using the FIFE Tools



NOvA: excludes maximal mixing at 2.5 s.d.



Dark Energy Survey: Optical follow-up of gravitational wave triggers



Microboone: ~~first~~ results

MINOS+: limits on LEDs