Neuroscience Gateway (NSG)

Developers Platform and Dissemination of Neuroscience Software

Subhashini Sivagnanam, Kenneth Yoshimoto, Martin Kandes, Steven Yeu, Amit Majumdar
San Diego Supercomputer Center, University of California San Diego

Ted Carnevale
Department of Neurobiology, Yale University

March 18, 2021

Outline

- **Recap of ongoing NSG activities and new features**
- Introduction to NSG software development platform
- “Tour” of the software development platform and functionalities
- Introduction to NSG software dissemination

- Feedback from distinguished invited attendees

The Neuroscience Gateway (NSG)

The NSG provides simple and secure access through portal and programmatic services, to run neuroscience modeling and data processing software on high performance and (more and more) high throughput compute resources

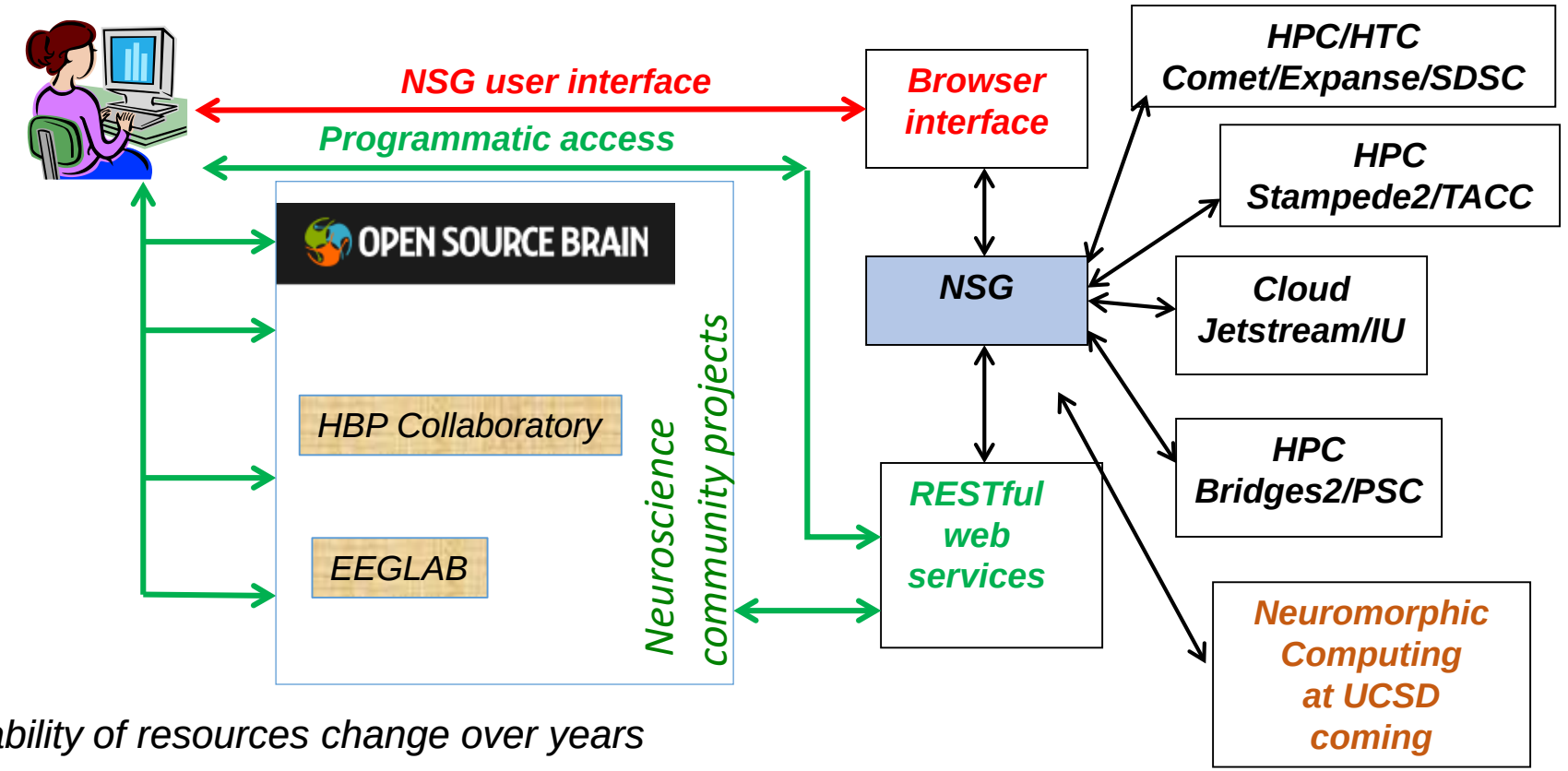
http://www.nsgportal.org

FREE and OPEN to any academic and non-profit researchers from any country

NSG catalyzes and democratizes computational and data processing neuroscience research and education for everybody including researchers and students from underrepresented institutions

NSG - Portal and Programmatic Access

- NSG Portal: Simple and easy to use web interface
- NSG–R: Programmatic access through RESTful services



Availability of resources change over years

Side note:
Lots of Supercomputer awards by NSF in 2019-2020-2021

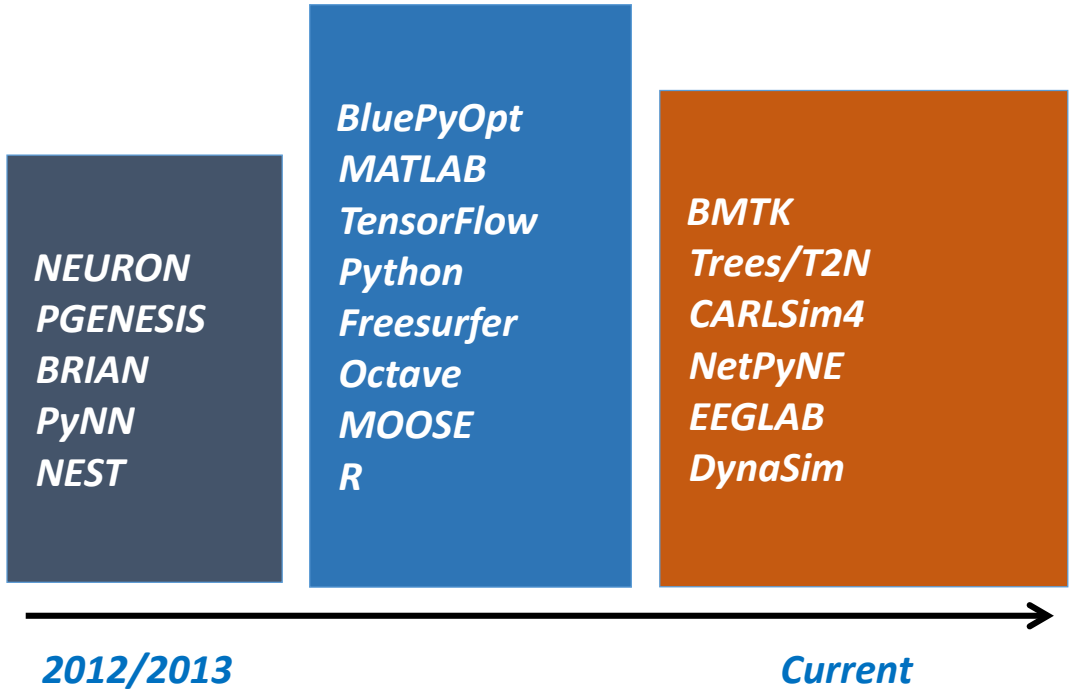
2020 December onwards in poduction:
- *Expanse at SDSC*
- *Bridges2 at PSC*

2021:
- *Anvil at Purdue*
- *Jetstream2 at IU*
- *Delta at NCSA, UIUC*

AI focused machines 2021:
- *Neocortex at PSC*
- *Voyager at SDSC*

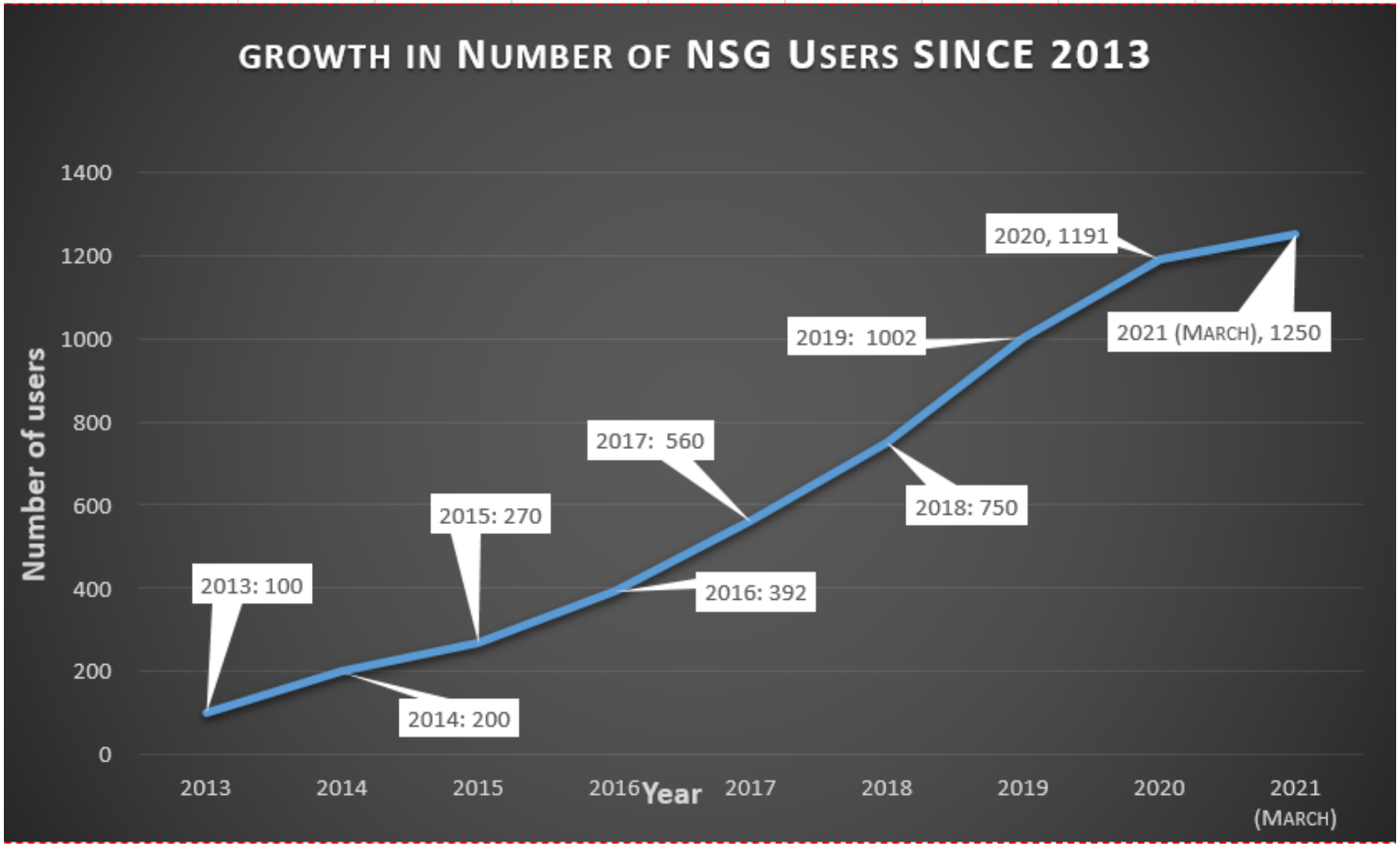
NSG modeling and data processing tools

(new tools added based on user needs and un-supported/un-used tools are retired)

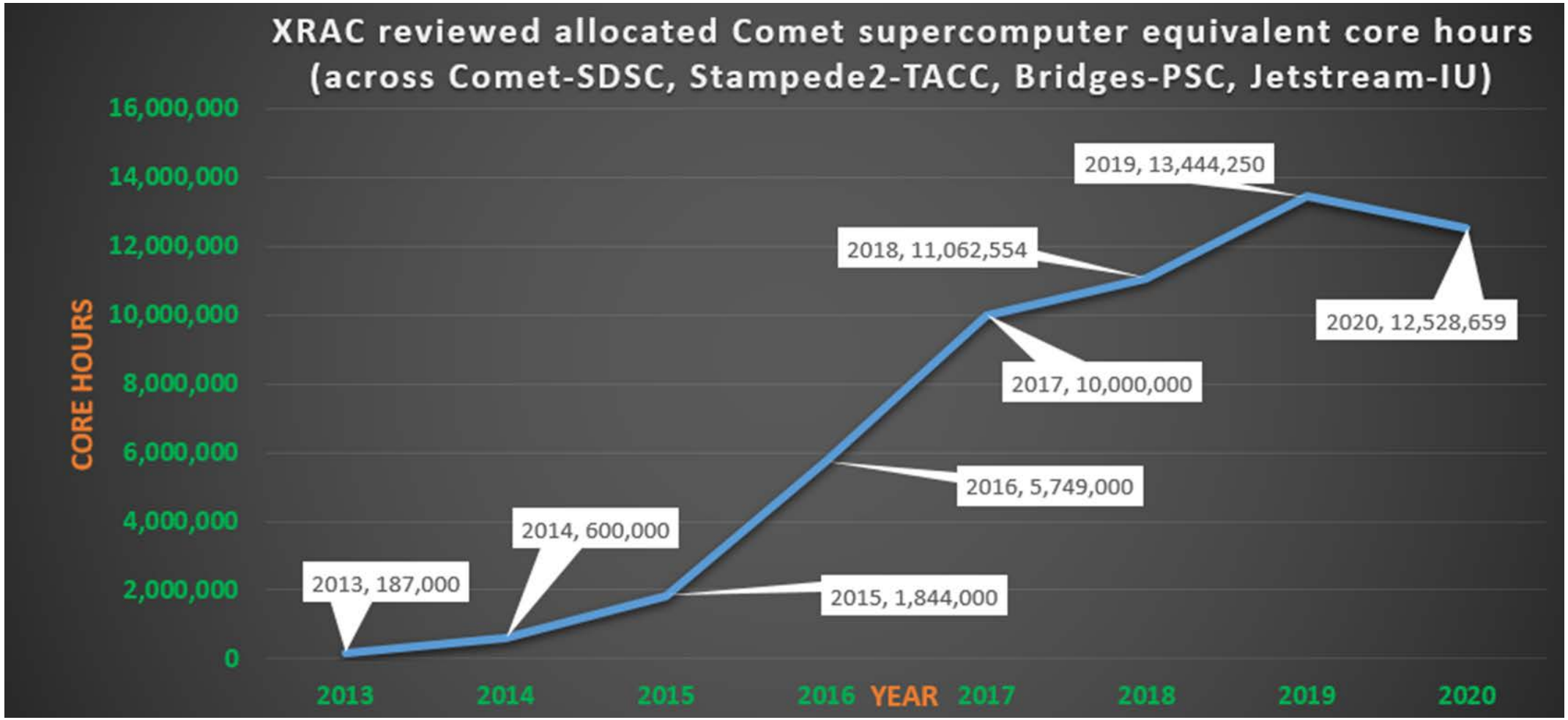


- **NSF funded Comet machine at SDSC will ramp down and start retirement starting April 2021**
- **New NSF funded supercomputer Expanse went into production starting December 2020**
  - **AMD EPYC 7742 processors and NVIDIA V100 SMX2 GPUs**
- **We are in the process of moving software to Expanse from Comet**

NSG Growth in number of registered users – since 2013



NSG Growth in supercomputer time allocated – since 2013



NSG Enabled Publications

- https://www.nsgportal.org/citation.htm

- Neuroscience publications, presentations, posters: **156** (that we know of)
- Cyberinfrastructure related publications, presentations, posters: **56**
- Educational projects/publications (MS/PhD thesis) and Training/workshops: **35**
- High school, undergrad student internship since **2011**

Comet (retiring) and Expanse (new) Supercomputers at SDSC

Comet System Configuration

System Component	Configuration
Intel Haswell Standard Compute Nodes	
Node count	1,944
Clock speed	2.5 GHz
Cores/node	24
DRAM/node	128 GB
SSD memory/node	320 GB
NVIDIA Kepler K80 GPU Nodes	
Node count	36
CPU cores:GPUs/node	24:4
CPU:GPU DRAM/node	128 GB:48 GB
NVIDIA Pascal P100 GPU Nodes	
Node count	36
CPU cores:GPUs/node	28:4
CPU:GPU DRAM/node	128 GB:64 GB
Large-memory Haswell Nodes	
Node count	4
Clock speed	2.2 GHz
Cores/node	64
DRAM/node	1.5 TB
SSD memory/node	400 GB
Storage Systems	
File systems	Lustre, NFS
Performance Storage	7.6 PB
Home file system	280 TB

5+ GB per core

- ***Comet will retire soon (ramp down rapidly starting April 2021) as a NSF Supercomputer***
- ***Expanse became available starting December 2020 for the next 5 years***
  - ***On Expanse NSG users need to specify memory requirement via NSG interfaces***

Expanse System Configuration

Full System	
Total compute nodes	728
Total compute cores	93,184
Total GPU nodes	52
Total V100 GPUs	208
Peak performance	5.16 PFlop/s
Total memory	247 TB
Total memory bandwidth	215 TB/s
Total flash memory	824 TB
HDR InfiniBand Interconnect	
Topology	Hybrid Fat-Tree

Compute Nodes	
CPU Type	*AMD EPYC 7742*
Nodes	726
Sockets	2
Cores/socket	64
Clock speed	2.25 GHz
Flop speed	4608 GFlop/s
Memory capacity	256 GB DDR4 DRAM
Local Storage	1TB Intel P4510 NVMe PCIe SSD
Max CPU Memory bandwidth	409.5 GB/s

~ <2 GB per core

GPU Nodes	
GPU Type	NVIDIA V100 SMX2
Nodes	52
GPUs/node	4
CPU Type	Xeon Gold 6248
Cores/socket	20
Sockets	2
Clock speed	2.5 GHz

Large-Memory	
CPU Type	*AMD EPYC 7742*
Nodes	4
Sockets	2
Cores/socket	64
Clock speed	2.25 GHz
Flop speed	4608 GFlop/s
Memory capacity	2 TB

DISK I/O Subsystem	
File Systems	NFS, Ceph
Lustre Storage(performance)	12 PB
Ceph Storage	7 PB

HDR InfiniBand Interconnect	
Topology	Hybrid Fat-Tree

NSG: evolving to meet the evolving needs of neuroscience research

- *Started in 2013 as a computational neuroscience science gateway – NEURON, GENESIS, NEST etc.*
- *In recent years more and more data processing software, pipelines – EEGLAB, Freesurfer, R etc.*

- *Big Data transfer, data sharing, data management were not part of NSG in 2013 or in the earlier years of NSG*
- *Due to growing data processing and analysis in neuroscience, those are a necessity for NSG to provide*

- *In recent years NSG is used for training, workshops, educational purposes from institutions worldwide*

- *In recent years increase in request for dissemination of neuroscience software*

New Features 1: NSG integration with Open Science Grid

- Open Science Grid – Pioneered High Throughput Computing (HTC); NSF, DOE funded
  - Provides software and services to users and resource providers
  - A distributed fabric of high throughput computational services
  - OSG does not own resources – uses available resources
  - Enables the opportunistic usage, sharing of resources
  - Uses meta-scheduling via HTCondor
  - OSG uses “glide in” mechanism to submit jobs to “free” resources at OSG resource providers
- NSG was integrated with OSG
  - NSG users see no difference jobs going to OSG resources
  - Typically limit on data size and single node jobs
  - First software made available on NSG portal is TensorFlow on OSG
  - In the future other neuroscience software can be made available
- Due to data size limit in the future explore StashCache architecture of OSG

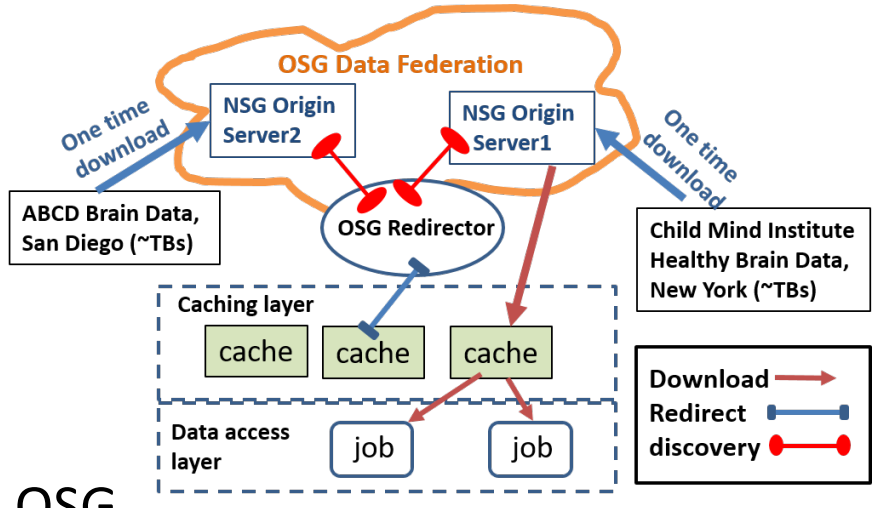


Figure 8. OSG StashCache Architecture.

New Features 2: large data transfer

- Until now, NSG allowed data upload via the portal, ok for small size files
- Data processing requires larger size files 10s/100s/1000s GB

- NSG was integrated recently with Globus data transfer
  - Reliable, high-performance file transfer platform
  - Allows users to transfer large amounts of data seamlessly between systems
  - Automatic transfers with built-in fault recovery

- Globus can be used for file transfers between the Neuroscience Gateway and:
  - An institution with Globus installed
  - A personal computer (known as a personal endpoint)
- NSG team provides information to users about directories into which data is transferred
- These directories can be accessed in user's job
- Community user mode of operation allows technically jobs to look into directories

New Features 3: data sharing

- Globus can also be used for sharing data with your collaborators
- Create a group of existing NSG users who can access your data
- Public and personal directory paths provided to users
- Data can be used by the user and user's collaborators in NSG jobs

Intended use of the directories	
Public directories	**Personal directories**
• To make data available to other NSG users to read and process • Not recommended for writing output files from NSG jobs	• Recommended for personal input data to be used in your own NSG jobs • Recommended for writing large output files from NSG jobs

Note: NSG operates using a community user account at the backend and technically it is possible for the community user account to access user's data from a job.

New Features 4: data access and management

- **Multiple users are now able to process same data** – publicly available data or data shared by a user
  - Other users can process the data using NSG's data processing software e.g. Matlab, Python, EEGLAB, Freesurfer etc.
  - This is outcome of the Globus sharing

- **We are able to provide users persistent storage**
  - A user can access same data from one job to another
  - For example output of one job can be accessed in the following job

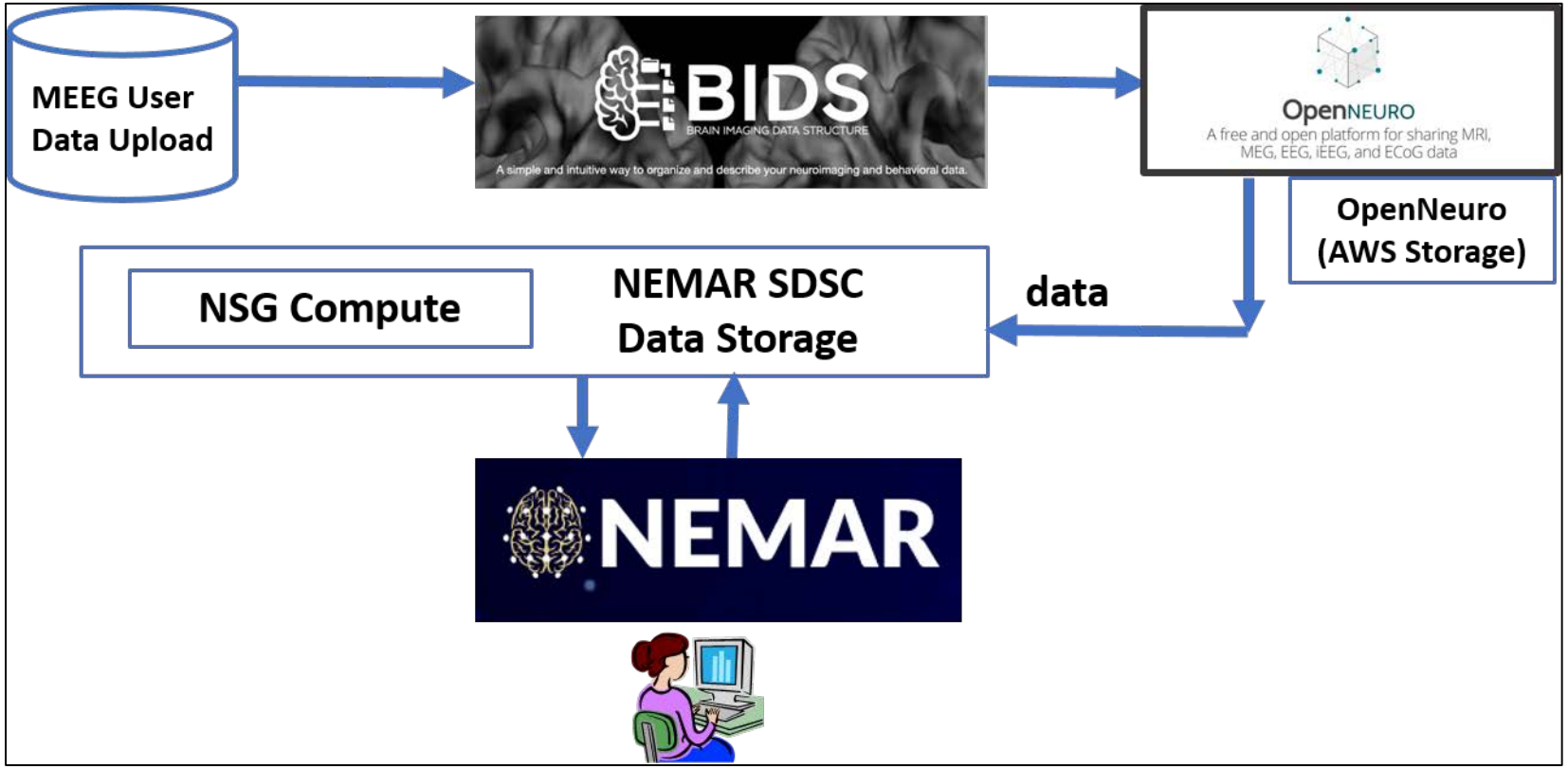
New Feature 5 (work in progress): commercial cloud integration

- First approach is cloudbursting from Expanse
- SDSC systems people are implementing cloudbursting (already on Comet) on Expanse
- A SLURM job can be tagged to send to AWS via direct path connection set up with AWS by network experts
- NSG integration ongoing
- If NSG has cloud $, we can send NSG jobs to AWS (and other cloud vendors eventually) via this mechanism
  - If individual users come with their own $ amount, cloud accounting not trivial to solve
- Integration with NSF funded CloudBank project for managing account
  - A cloud access entity that helps access and use public of clouds for research and education by delivering a set of managed services
  - https://www.cloudbank.org/

Training and Outreach

- NSG training/workshop regularly since 2013: at SfN, CNS, CogSci conferences, other
- Last training webinar ~90 attendees
  - **Neuroscience Gateway Using NEURON and EEGLAB - Training Webinar,** Friday, November 20, 2020
- Upcoming training in June 2021
  - Specifically for HSI neuroscience faculties
  - Tentative title: *Introduction to Neuroscience Gateway and Electroencephalography, EEGLAB*
- NSG used in classroom teaching
- Continuing since 2011 Research Experience for High School (REHS) program
  - 2020 project topics included HPC and Mobile EEG device collected data analyzed using EEGLAB

NSG part of the NIH BRAIN Initiative NEMAR grant
NeuroElectroMagnetic data Archive and tools Resource



- *Develop NEM data archive and data curation workflow*
- *Support both standard and custom data processing pipeline*
- *Support display, viz, download of data*
- *Support search of data*
- *Use NSG for data processing*
- *Connected to OpenNeuro*

Outline

- Recap of ongoing NSG activities and new features
- **Introduction to NSG software development platform**
- "Tour" of the software development platform and functionalities
- Introduction to NSG software dissemination

- Feedback from distinguished invited attendees

Observations from NSG usage mode in recent years

- Developers of neuroscience modeling and data processing software are **using NSG as a cyberinfrastructure environment on which they can do development, testing and benchmarking (including at-scale) of their software, libraries and pipelines**

- **Neuroscientists disseminating their developed software via NSG**

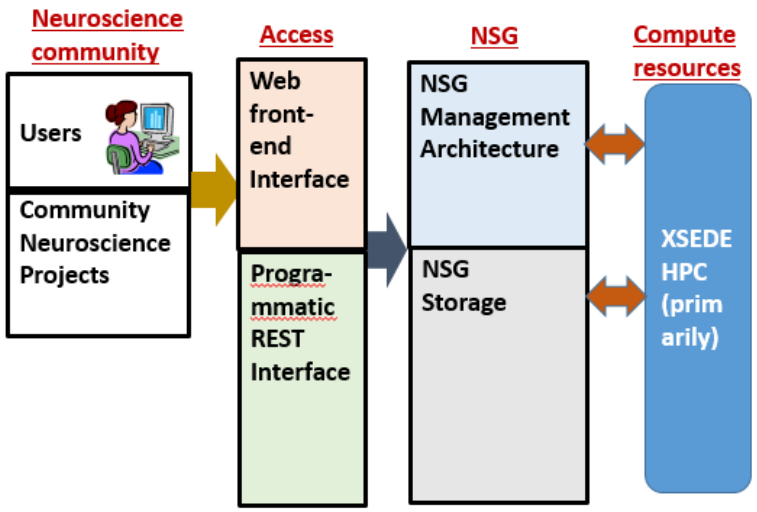
- **Motivation behind disseminating software via NSG:**

  - NSG already provides access to HPC and HTC computing resources, and GPUs
  - NSG provides optimally installed neuroscience and other needed software available on NSG's compute resources
  - NSG has a growing user base
  - NSG team works closely with neuroscience software developers
  - NSG has good user support mechanism
  - NSG has good training and outreach programs; hosts workshops jointly with developers and users
  - NSG can be mentioned in grants as one of the dissemination platforms

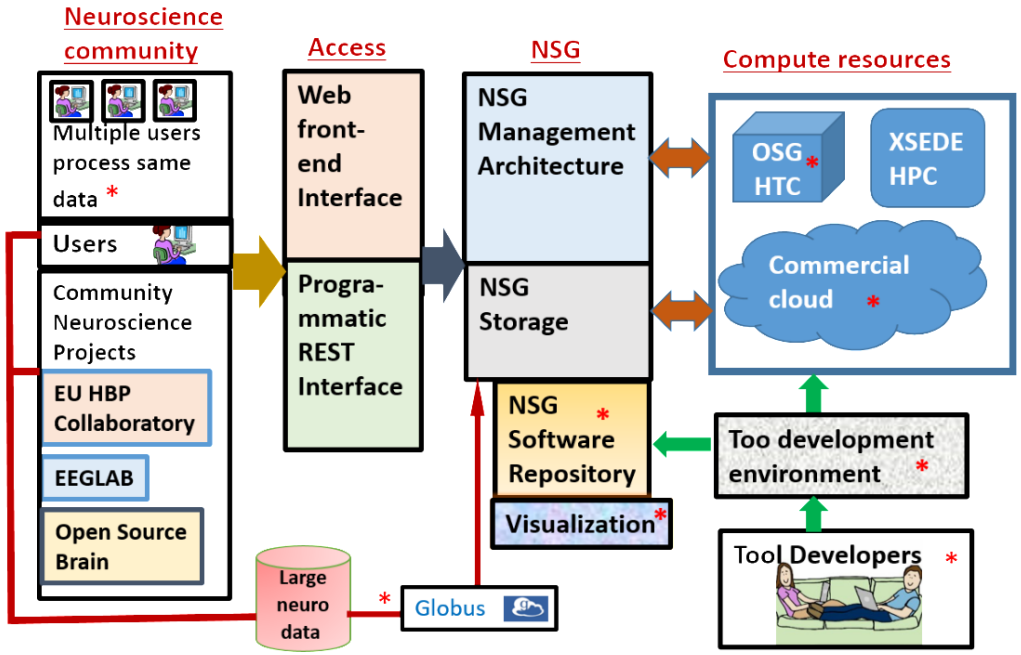
  - (BRAIN INITIATIVE Tools call encourages disseminating via archives such as NEMAR)

NSG – existing science gateway
Adding software development platform and dissemination capability

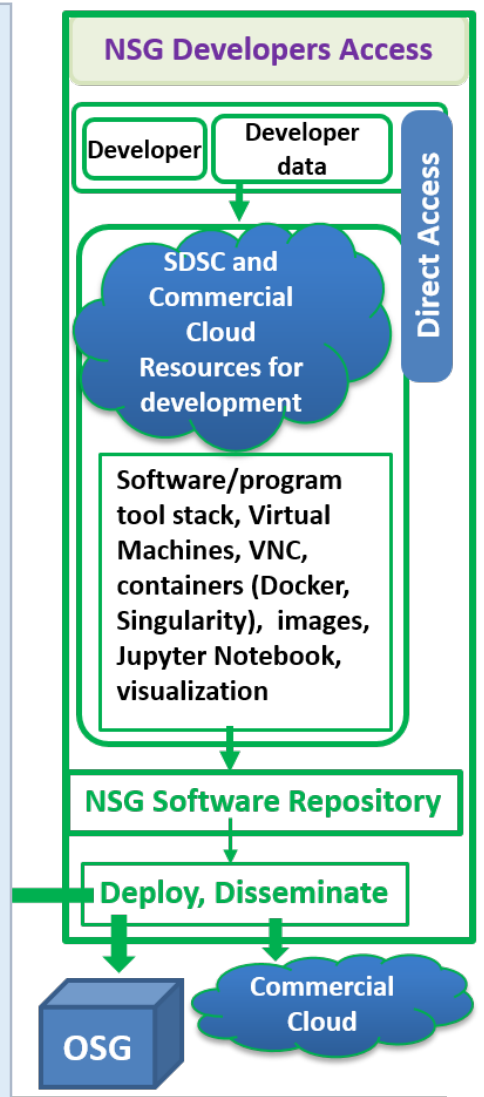
Original NSG



Evolving NSG
* *New features*

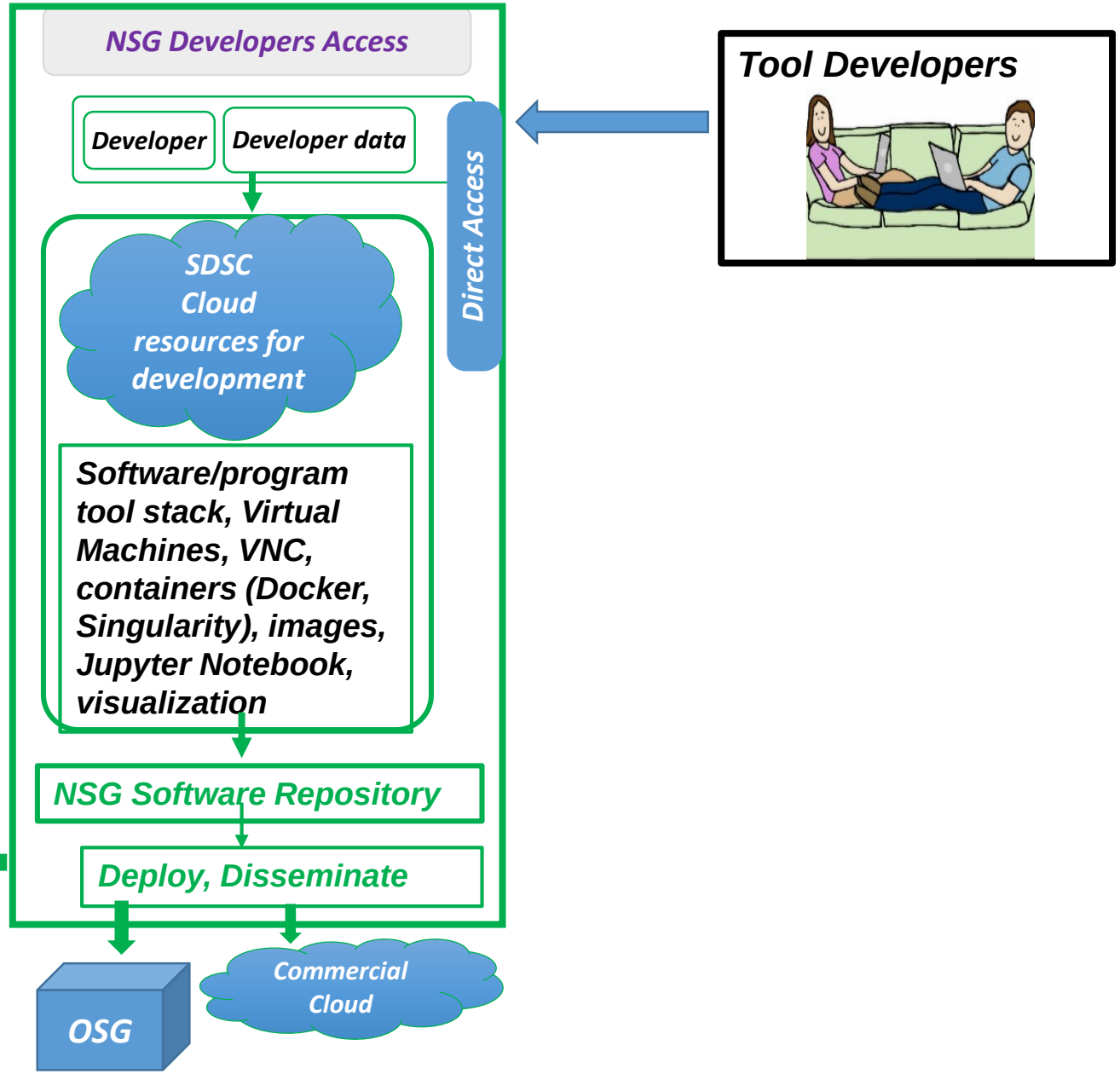


Addition of software development platform to current NSG



(b) New Development, Dissemination Component.

Various components of software development platform



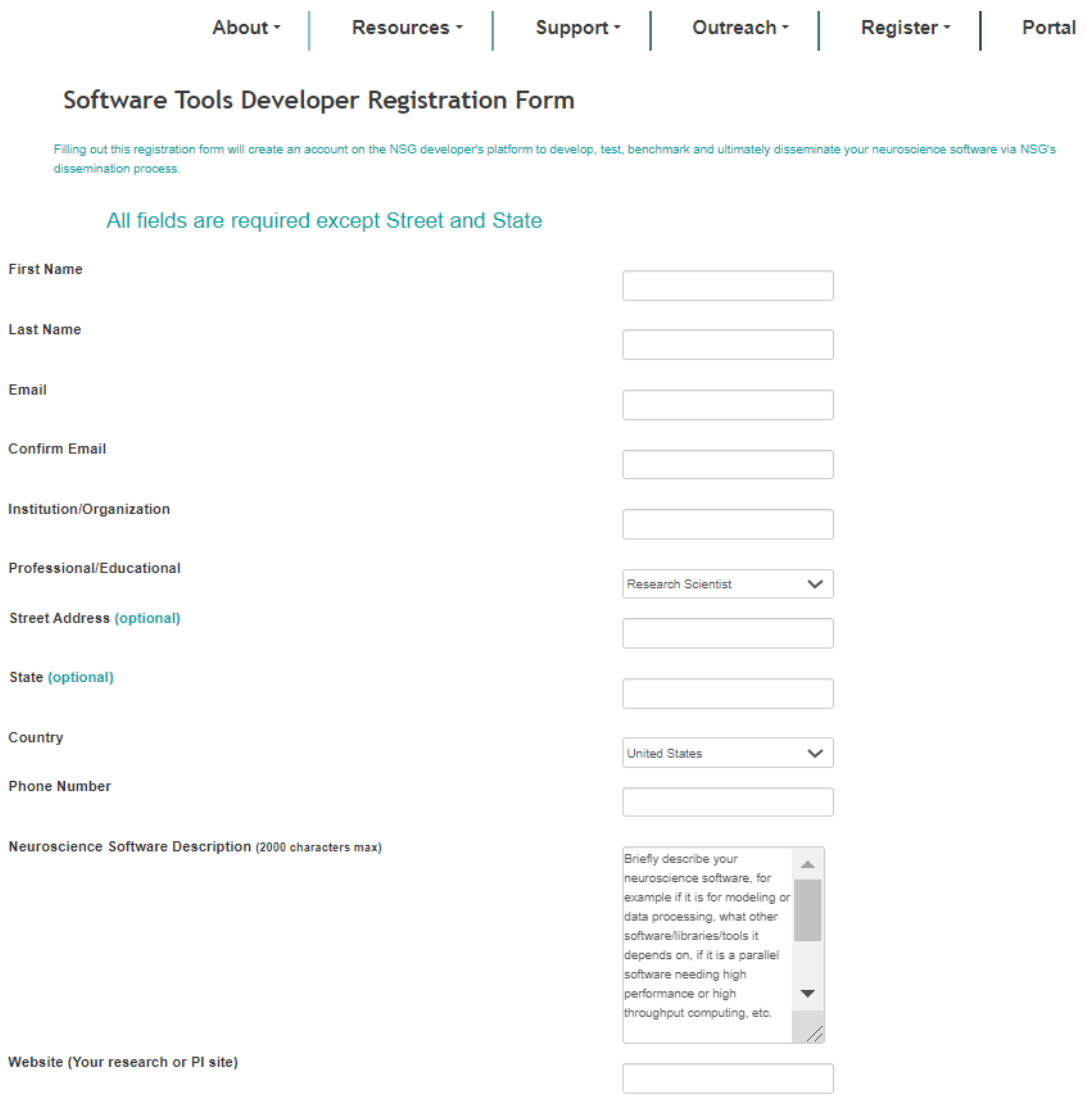
Software Development Platform

- Provide neuroscience software developers direct access to a SDSC cloud platform (free to developers)
- Neuroscience software
  - development
  - enhancement
  - testing
  - benchmarking of modeling and data processing neuroscience software
- In addition to SDSC cloud, access to HPC/HTC, GPU computing resources
- Access to commercial cloud computing for development, testing
- Test implementation on production environment of job submission (via scheduler), file system, data access etc.

Enabling neuroscience software development

- Enable development of neuroscience software that are implemented
  - in shared memory (such as threaded codes) parallel across the cores of a multi-core node
  - in distributed memory (such as using the Message Passing Interface library) parallel
  - Hybrid parallel (MPI-OpenMP)
  - Accelerators (GPUs; in the future other special AI hardware)
- The SDSC cloud platform contains a comprehensive software stack
  - OS (such as Ubuntu, CentOS)
  - Compilers (such as Intel, gcc)
  - Middleware (such as MPI library)
  - Analysis software and packages (such as Matlab, R, Python)
  - computational and data processing neuroscience software (NEURON, EEGLAB, ML/DL tools etc.)
  - collaborative tools (such as VNC server, Jupyter Notebook)
  - Container (Docker and Singularity containers)
- Additional software can be added based on neuroscience software developers' needs

Register as a developer to get access to NSG's SDSC cloud environment



Outline

- Recap of ongoing NSG activities and new features
- Introduction to NSG software development platform
- **"Tour" of the software development platform and functionalities**
- Introduction to NSG software dissemination

- Feedback from distinguished invited attendees

Outline

- Recap of ongoing NSG activities and new features
- Introduction to NSG software development platform
- “Tour” of the software development platform and functionalities
- **Introduction to NSG software dissemination**

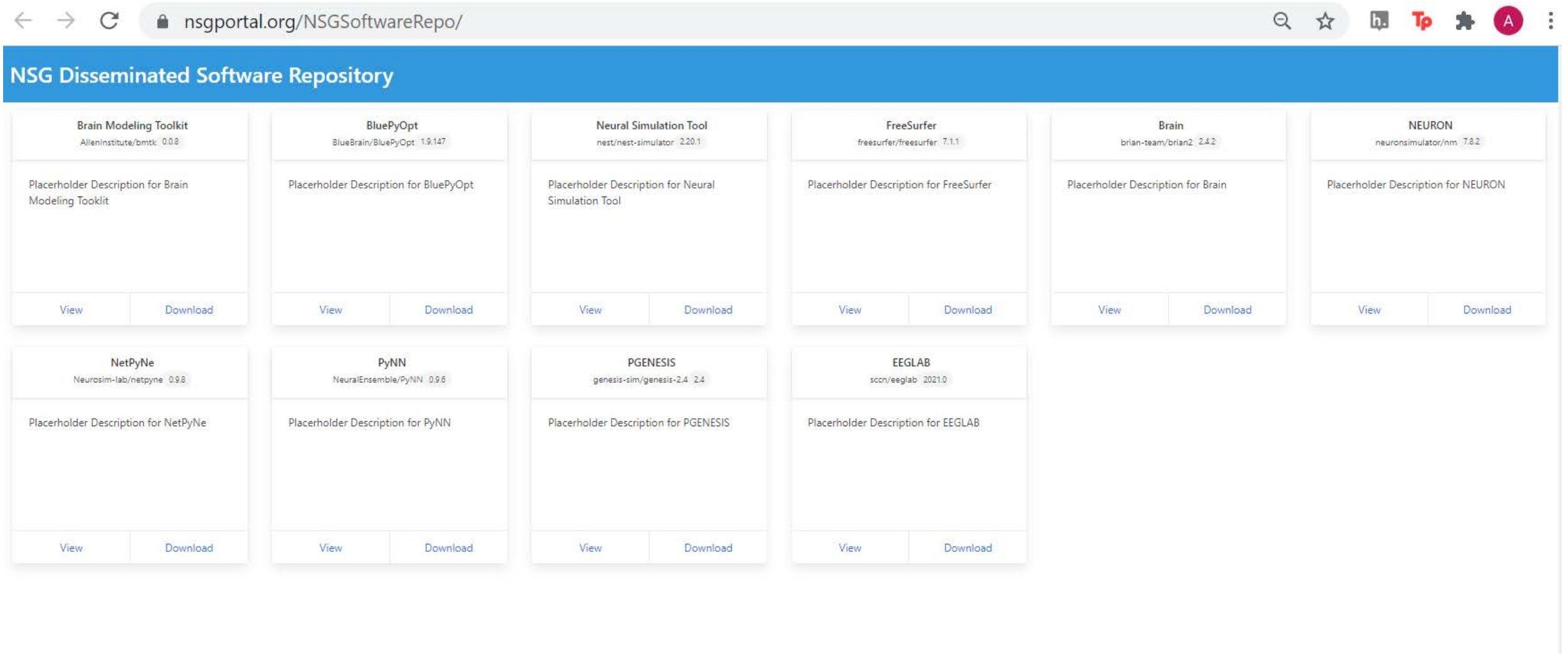
- Feedback from distinguished invited attendees

NSG neuroscience software repository

- NSG will provide software repository for mature and robust neuroscience software
- NSG software repository on SDSC's cloud storage – TBs of space (paid by NSG grant)
  - Storage free for developers disseminating software
- Developers will be able to move software to NSG repository from
  - NSG software development platform (fast)
  - Outside development environment

- A web frontend lists all the software disseminated
  - Provides detailed description of the software for the user community
  - Points to software's GitHub location
  - Provides information about input/output files (especially related to NSG)

- Web frontend allows users to download the software – singularity image, cloud image etc.


nsgportal.org/NSGSoftwareRepo/
NSG Disseminated Software Repository
Brain Modeling Toolkit
AllenInstitute/bmtk 0.0.8
Placerholder Description for Brain Modeling Tooklit
View
Download
BluePyOpt
BlueBrain/BluePyOpt 1.9.147
Placerholder Description for BluePyOpt
View
Download
Neural Simulation Tool
nest/nest-simulator 2.20.1
Placerholder Description for Neural Simulation Tool
View
Download
FreeSurfer
freesurfer/freesurfer 7.1.1
Placerholder Description for FreeSurfer
View
Download
Brain
brian-team/brian2 2.4.2
Placerholder Description for Brain
View
Download
NEURON
neuronsimulator/nrn 7.8.2
Placerholder Description for NEURON
View
Download
NetPyNe
Neurosim-lab/netpyne 0.9.8
Placerholder Description for NetPyNe
View
Download
PyNN
NeuralEnsemble/PyNN 0.9.6
Placerholder Description for PyNN
View
Download
PGENESIS
genesis-sim/genesis-2.4 2.4
Placerholder Description for PGENESIS
View
Download
EEGLAB
sccn/eeglab 2021.0
Placerholder Description for EEGLAB
View
Download

Neuroscience software dissemination

- Developers who use NSG developer platform can disseminate their software by publicly sharing their software and guidelines for usage via NSG
  - Software can be installed on variety large scale free XSEDE compute resources
  - HPC/HTC/cloud/GPU resources available at SDSC, TACC, PSC, other supercomputer centers

- Software can be made available in a containerized form, a cloud image etc.
  - Software can be downloaded by interested researchers
  - Users can use software on other computing resources they have access to
  - Users can take cloud image to commercial cloud resources (and pay from their own funding)

- To avoid “person in the middle” situation
  - We will provide software developers access to NSG’s SDSC file system where software resides
  - Developers can maintain, update software for SDSC HPC/HTC resources
  - Developers can set up cron jobs to update software

NSG: outreach and training

- NSG interacts with NeuroImaging Tools and Resources Collaboratory [NITRC] for dissemination
  - Can enable distribution via NITRC

- NSG disseminated software available to 1250+ NSG users – researchers, students

- Train users and a natural outreach for software at NSG workshops

- Software can be used by students in classroom teaching

NSG - Summary

- NSG is FREE and OPEN to academic and non-profit researchers
- Provides programmatic and portal access to neuroscience tools, pipelines, data processing software on HPC, HTC, GPU and academic cloud resources ; commercial cloud work on going
- NSG works with developers for neuroscience software

- **Adding new features of software development and dissemination capability to existing NSG**

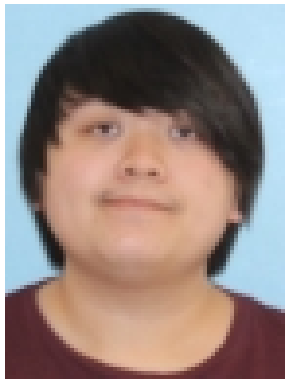
- Encourage collaboration with researchers from around the world; strongly encourage high school, undergraduate student participation in summer projects with NSG.
- Workshops at Society for Neuroscience annual meetings, Computational Neuroscience annual meetings, NEURON Summer Workshops, CogSci18, MSI institutions (New Mexico State University, Cal State San Bernardino), Neuroinformatics, Janelia Research Campus etc.
- Please cite us if you use NSG (http://www.nsgportal.org/citation.html) S Sivagnanam, A Majumdar, K Yoshimoto, V Astakhov, A Bandrowski, M. E. Martone, and N. T. Carnevale. Introducing the Neuroscience Gateway, IWSG, volume 993 of CEUR Workshop Proceedings, CEUR-WS.org, 2013
- If NSG was used, please let us know of your talks, presentations, publications, thesis work so that we can include in reports – nsghelp@sdsc.edu

NSG Team

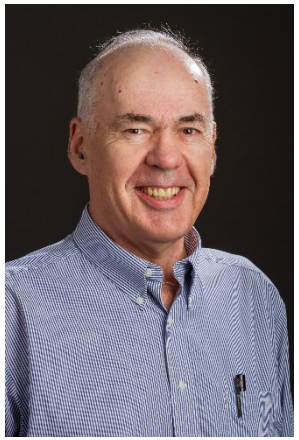
Subha, UCSD

Kenneth , UCSD

Marty, UCSD



Steven, UCSD



Ted, Yale U.

Amit, UCSD

NSG Collaborators

Many collaborators:

- M. Hines, R. McDougal, T. Morse, Yale U; B. Lytton, S. Dura-Bernal, SUNY DMC; S. Makeig, A. Delorme, G. Cauwenberghs, M. Miller, A. Bandrowski, M. Martone, M. Tatineni UCSD; S. Nair, U. Missouri; M. Migliore, C. Lupascu IRC Italy; P. Kumbhar, HBP, EPFL; S. Neymotin Nathan S. Kline Inst; A. Silver, P. Gleeson, UCL London; D. Beeman, U Colorado; U. Bhalla, NCBS-India; J. Krichmar, UCI
- PIs/developers/supporters of software (BluePyOpt, BMTK, Brian, DynaSim, Freesurfer, MOOSE, NEST, PyNN etc.)

NSG Advisory Board:

- Maxim Bazhenov, UCSD
- Dieter Jaeger, Emory University
- Stephanie Jones, Brown University
- William Lytton, SUNY Downstate Medical Center
- Vinod Menon, Stanford University

Acknowledgement of grants

- NIH NIBIB grant: Neuroscience Gateway to Enable Dissemination of Computational and Data Processing Tools and Software (NIBIB IU24EB029005; UCSD, Yale; 2019 - 2024)
- NSF DBI grant: Collaborative Research: CIBR: Building Capacity for Data-driven Neuroscience Research (#1935749 UCSD, #1935771 Yale University; 2020 – 2023)
- NIH NIMH grant: BRAIN INITIATIVE RESOURCE: Development of a Human Neuroelectromagnetic Data Archive and Tools Resource (NEMAR) (NIMH R24MH120037; UCSD, Stanford; 2019 – 2024)
- NSF CNS grant: CRI: CI-NEW: Trainable Reconfigurable Development Platform for Large-Scale Neuromorphic Cognitive Computing (#1823366, UCSD, UCI; 2018 – 2021)
- NIH NIBIB grant: The Open EEGLAB Portal Project (NIBIB R01EB023297; UCSD; 2017 – 2021)
- NSF OAC grant: CyberTraining: DSE: Self-Service Training Modules for Data-Intensive Neuroscience Learning and Research (#1730655; U. Missouri, UCSD; 2017 – 2021)

Feedback from distinguished attendees