The OSG Fabric of Services and Cloud Resources

Miron Livny
John P. Morgridge Professor of Computer Science
OSG Technical Director
UW-Madison/CHTC

PATh PARTNERSHIP to ADVANCE THROUGHPUT COMPUTING

Open Science Grid

HT CENTER FOR HIGH THROUGHPUT COMPUTING

Terminology

- Will use “Cloud” to refer to **Commercial** Clouds, **OpenStack** environments and **Kubernetes** (k8s) clusters
- Will use “resources” to refer to CPUs, GPUs, main Memory and attached/local storage

Shared Storage can be a topic for a future presentation

How do Cloud Resources fit into the Distributed High Throughput Computing (dHTC) model of the fabric of Services provided by the Open Science Grid (OSG)?

PATh PARTNERSHIP to ADVANCE THROUGHPUT COMPUTING

HT CENTER FOR HIGH THROUGHPUT COMPUTING

Naturally!

It is all about offering
Access Points
and services to deploy
Execution Points

PATh PARTNERSHIP to ADVANCE THROUGHPUT COMPUTING

HT CENTER FOR HIGH THROUGHPUT COMPUTING

CERN 92

(CERN, is the world's largest particle physics laboratory)

Flock

Global Scientific Computing via a Flock of Condors

Miron Livny

Computer Sciences Department
University of Wisconsin — Madison
Madison, Wisconsin
{miron@cs.wisc.edu}

Flock

MISSION

Give scientists effective and efficient access to large amounts of cheap (if possible free) CPU cycles and main memory storage

Flock

THE CHALLENGE

How to turn existing privetly owned clusters of *workstations*, *farms*, *multiprocessors*, and *supercomputers* into an efficient and effective Global Computing Environment?

In other words, how to minimize wait while idle?

Flock

APPROACH

Use wide-area networks to transfer batch jobs between Condor systems

- Boundaries of each Condor system will be determined by physical or administrative considerations

Flock

TWO EFFORTS

☐ ***UW CAMPUS***

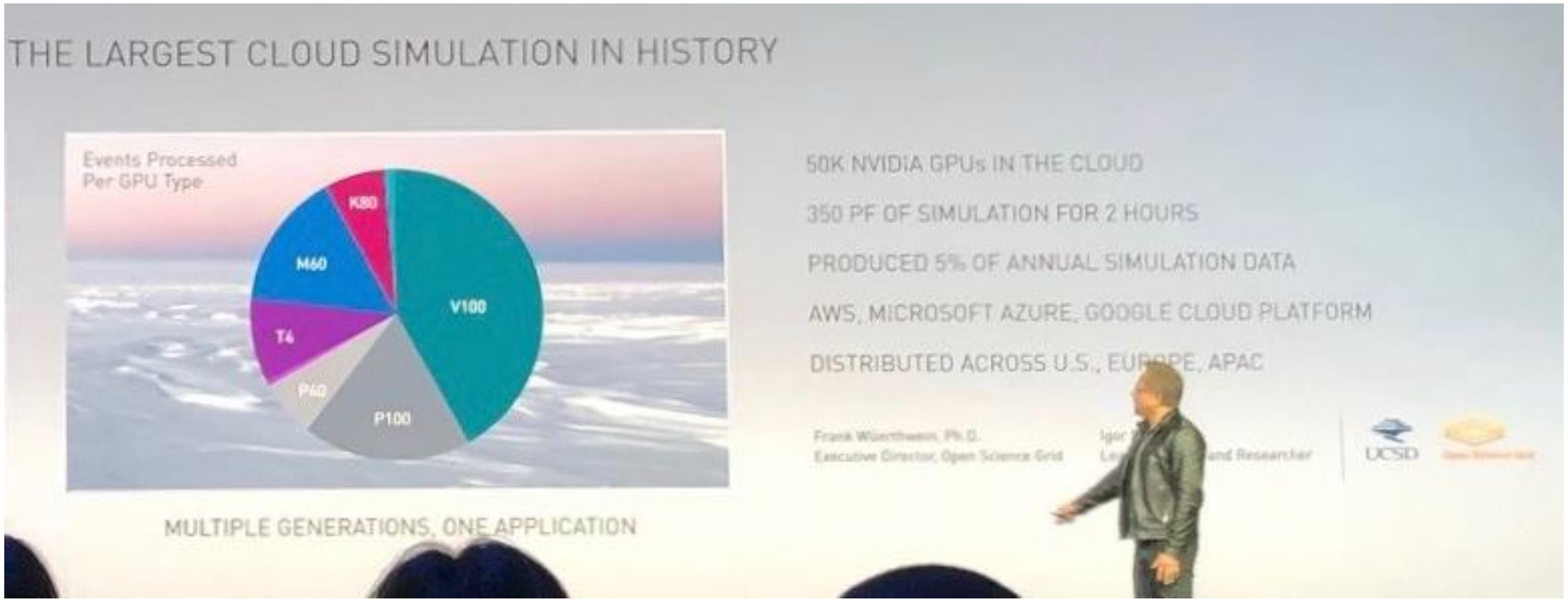
Condor systems at Engineering, Statistics, and Computer Sciences

☐ ***INTERNATIONAL***

We have started a collaboration between CERN-SMC-NIKHEF-Univ. of Amsterdam, and University of Wisconsin-Madison

Jensen Huang keynote @ SC19

The Largest Cloud Simulation in History



Saturday morning before SC19 we bought all GPU capacity that was for sale in Amazon Web Services, Microsoft Azure, and Google Cloud Platform worldwide

OSG Consortium

- Established in 2005, the Open Science Grid is a consortium of **stakeholders** governed by a **council**
- Members include campuses, research collaborations, software providers and compute, storage, networking providers
- The OSG provides a **fabric of dHTC Services** to the consortium members and to the broader Science and Engineering (S&E) community
- While **members own** and operate resources, the consortium does not own or operate any resources
- Council elects the OSG **Executive Director** who appoints an **Executive team**. Together they steer and manage available effort

PATh PARTNERSHIP to ADVANCE THROUGHPUT COMPUTING

Open Science Grid

HT CENTER FOR HIGH THROUGHPUT COMPUTING

OSG Fabric of Services

- Organized under three main thrusts – *Community Building, Research Computing Facilitation, and Operation*
- The consortium claims that its services enabled in the past **12** month more than **2B** core hours across more than **130** clusters located at more than **70** sites and more than **200TB** of data cached across **17** caches worldwide, sized at **30-100TB** each.
- Past Funding - **2006-2011** NSF (MPS) and DoE, **2012-2015** NSF (MPS,OAC) and DoE, **2016-2019** NSF (MPS, OAC)

PATh PARTNERSHIP to ADVANCE THROUGHPUT COMPUTING

Open Science Grid

HT CENTER FOR HIGH THROUGHPUT COMPUTING

Current Effort

- 66% of effort provided by the **Partnership to Advance Throughput Computing (PATh)** (NSF-OAC - 2030508)
- 22% of effort provided by the **S2I2: Institute for Research and Innovation in Software for High Energy Physics (IRIS-HEP)** (NSF-OAC 1836650)
- ~10% of effort provided by other projects and organizations

Aligned with the NSF blueprint for Cyberinfrastructure Coordination Services

PATh PARTNERSHIP to ADVANCE THROUGHPUT COMPUTING

Open Science Grid

HT CENTER FOR HIGH THROUGHPUT COMPUTING

Manish Parashar

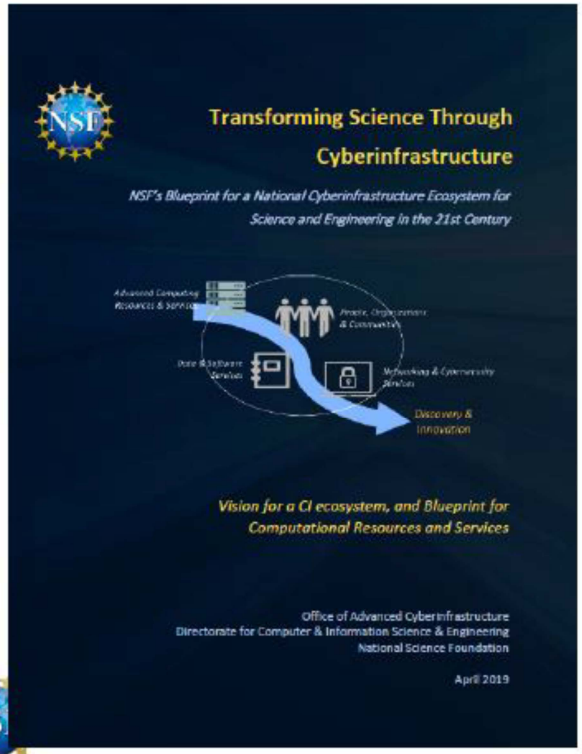
Director, Office of Advanced Cyberinfrastructure,

Directorate for Computer & Information Science & Engineering

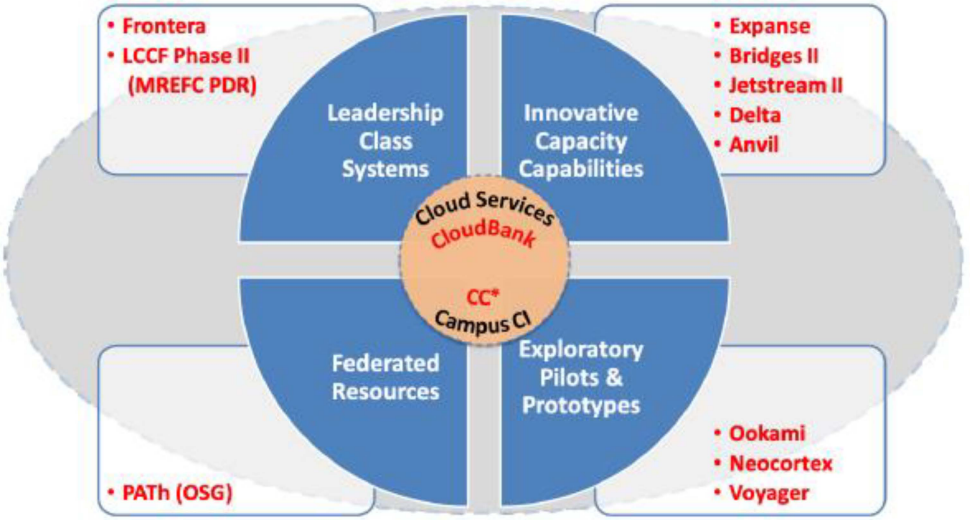
National Science Foundation

November 30, 2020

NSF's Vision for a National CI Ecosystem



Computational Ecosystem: Elements



PATh PARTNERSHIP to ADVANCE THROUGHPUT COMPUTING

Open Science Grid

HT CENTER FOR HIGH THROUGHPUT COMPUTING

The Open Science Pool (OSPool)

One of the OSG services are Access Points (**APoint**) that are open to any US researcher and a distributed HTCondor pool that is managed under a fair-share scheduling policy

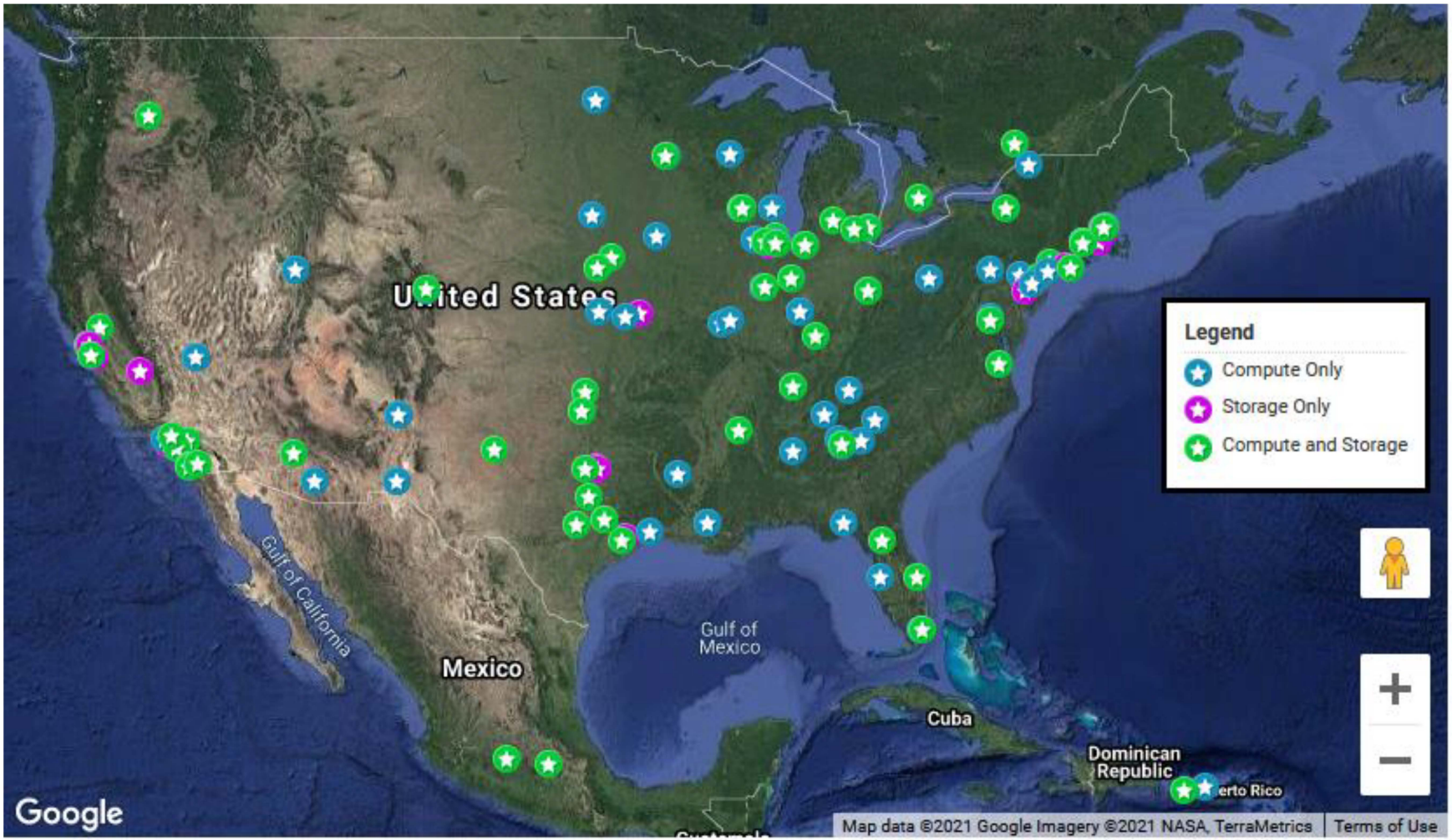
OSG Compute Federation sites contribute resources to the OSPool by (dynamically) running Execution Points (**XPoints**) according to autonomous site policies

- APoint provides workload automation and workflow management (DAGMan, Pegasus) capabilities designed to accommodate High Throughput applications
- Data for input sandboxes is staged at the APoint or placed in the OSG data federation
- Output sandbox data is staged at the APoint
- Yesterday (04/06/2021), the OSPool completed ~135K jobs from 36 projects that consumed ~900K core hours

PATh PARTNERSHIP to ADVANCE THROUGHPUT COMPUTING

Open Science Grid

HT CENTER FOR HIGH THROUGHPUT COMPUTING



PATh PARTNERSHIP to ADVANCE THROUGHPUT COMPUTING

Open Science Grid

HT CENTER FOR HIGH THROUGHPUT COMPUTING

Democratizing Access

In her presentation at the NSFNET 35th Anniversary NSF CISE AD Margert Martonosi articulated the challenge of **Democratizing Access to National Research Computing Resources**.

We view Access Points as holding the key to addressing this national challenge

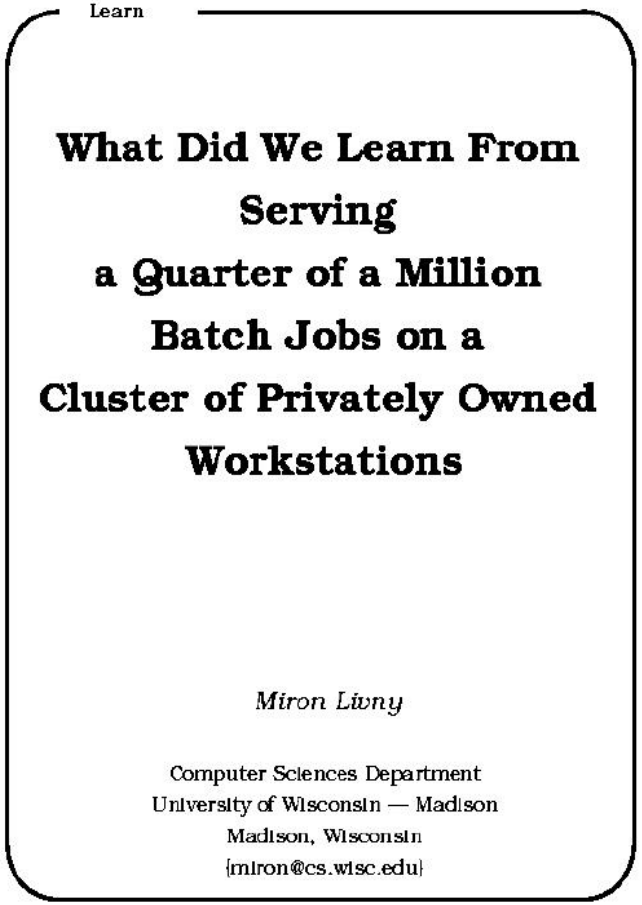
- OSPool APoints can be deployed and operated by a single PI laboratory or by organizations like campuses and science collaborations
- APoints can be used to manage deployment of XPoints
- APoints can scale out to accommodate large HTC workloads

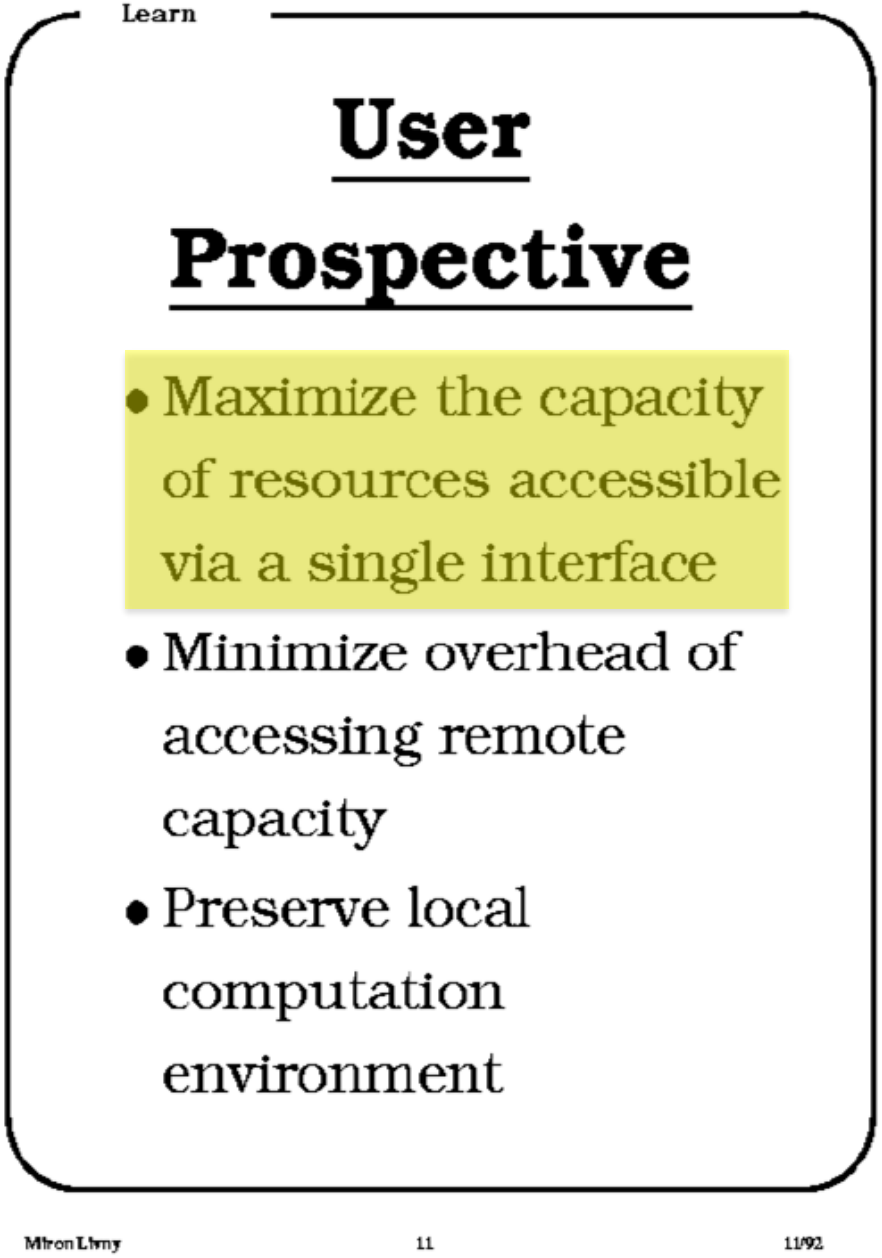
PATh PARTNERSHIP to ADVANCE THROUGHPUT COMPUTING

Open Science Grid

HT CENTER FOR HIGH THROUGHPUT COMPUTING

Miron Livny and Michael Litzkow, "Making Workstations a Friendly Environment for Batch Jobs", *Third IEEE Workshop on Workstation Operating Systems*, April **1992**.





Led to the formulation of the principal

Submit locally

(*store and manage your resource acquisition and job execution locally*)

Run globally

(*acquire and use any resource that is capable and willing to run your workload*)

PATh PARTNERSHIP to ADVANCE THROUGHPUT COMPUTING

Open Science Grid

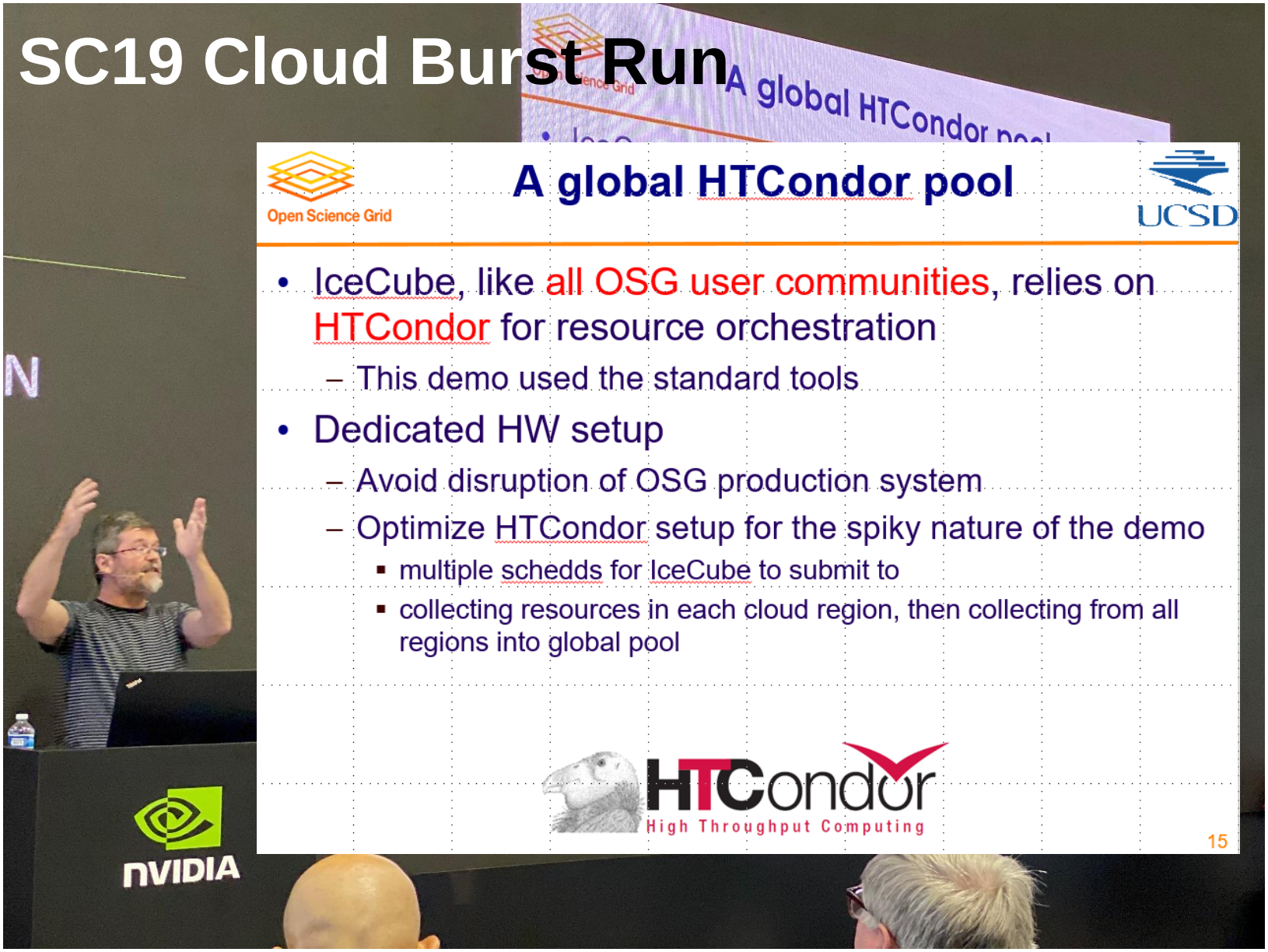
HT CENTER FOR HIGH THROUGHPUT COMPUTING

How do sites contribute to the OSPool?

When a site wants to contribute the capacity of a server to the OSPool, it runs an XPoint on the server

- Site can start and stop the XPoint at any time
- XPoint needs to establish trust with APoint
- OSG provides services to remotely activate XPoints through a Compute EntryPoint (CE) that submits activation requests to the batch system of the cluster
- OSG provides services to automate remote activation of XPoints
- XPoint prefers to have out going network connectivity

SC19 Cloud Burst Run

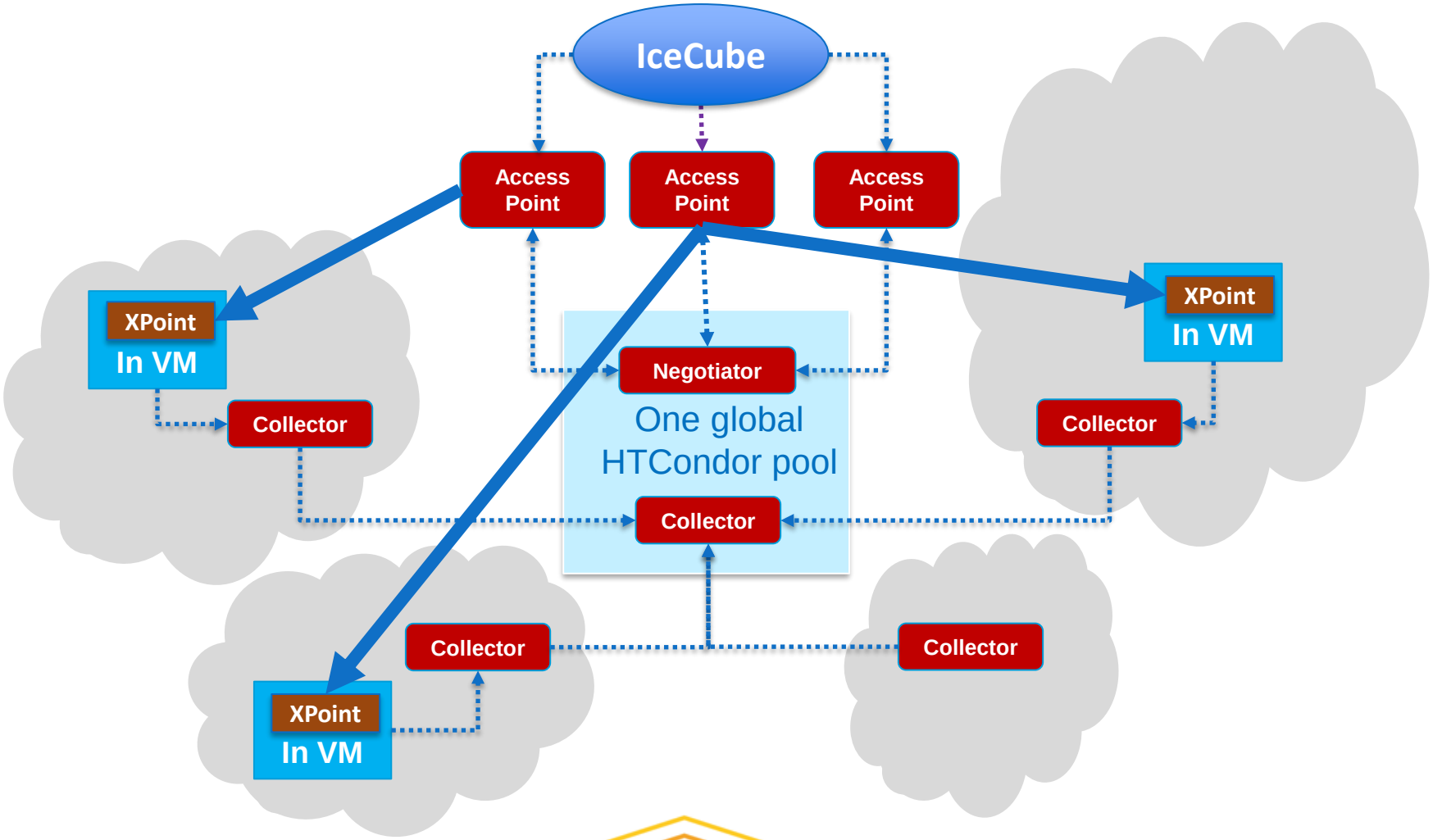


PATh PARTNERSHIP to ADVANCE THROUGHPUT COMPUTING

Open Science Grid

HT CENTER FOR HIGH THROUGHPUT COMPUTING

SC19 HTCondor Distributed CI



PATh PARTNERSHIP to ADVANCE THROUGHPUT COMPUTING

Open Science Grid

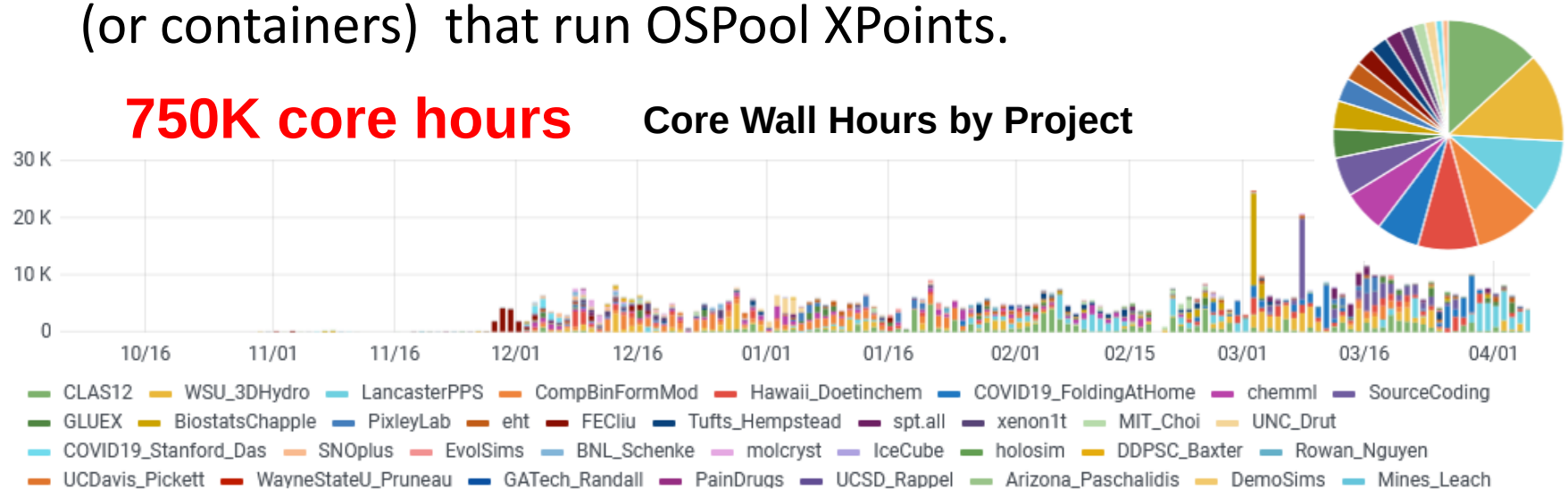
HT CENTER FOR HIGH THROUGHPUT COMPUTING

OpenStack Environment

The **Jetstream** cloud is an OpenStack, NSF-funded academic cloud operated by the Indiana University and TACC.

Operators decide when to start and terminate Virtual Machines (or containers) that run OSPool XPoints.

750K core hours



Work with NSF funded **Chameleon** is progressing nicely

PATh PARTNERSHIP to ADVANCE THROUGHPUT COMPUTING

Open Science Grid

HT CENTER FOR HIGH THROUGHPUT COMPUTING

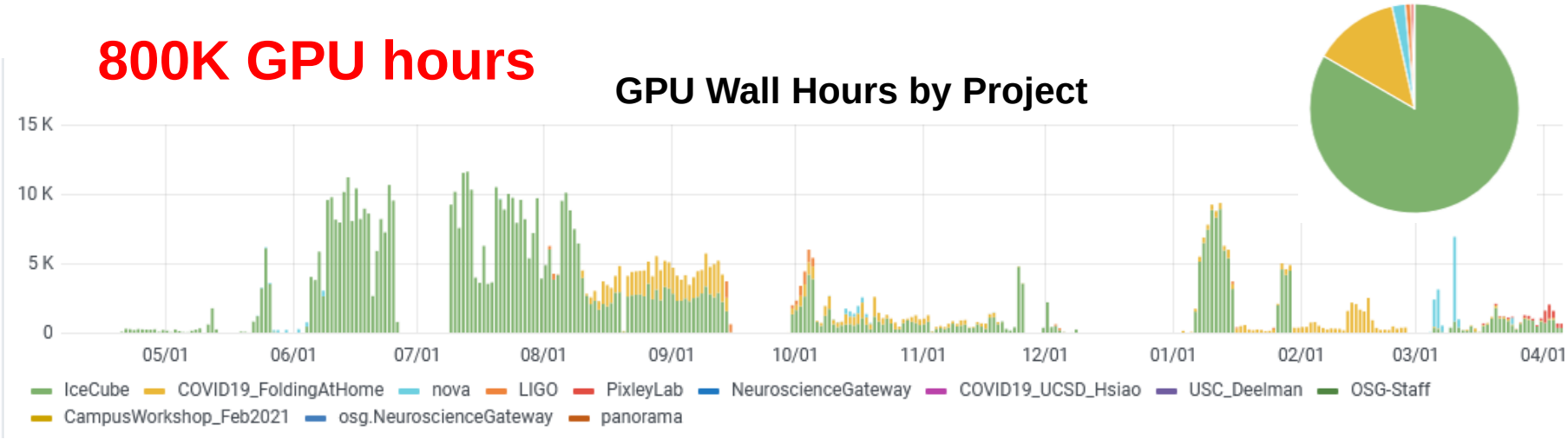
k8s Cluster

The NSF funded Pacific Research Platform **PRP** operates a k8s cluster across more than 35 location worldwide.

A PRP HTCondor pool grows and shrinks under the control of the k8S cluster pod scheduler

A CE submits requests to HTCondor to deploy OSPool XPoints.

800K GPU hours



PATh PARTNERSHIP to ADVANCE THROUGHPUT COMPUTING

Open Science Grid

HT CENTER FOR HIGH THROUGHPUT COMPUTING

OSPool is not the only pool!

Organizations like science collaborations (CMS, LIGO, IceCube) and campuses (UCSD,UNL, UW-Madison, JLab) leverage OSG services to deploy and operate private distributed HTCondor pools

- APoints are private
- Deployment of XPoints negotiated by the organization with sites
- Resource acquisition and allocation policies defined by the organization

* A site can contribute to different HTCondor pools

PATh PARTNERSHIP to ADVANCE THROUGHPUT COMPUTING

Open Science Grid

HT CENTER FOR HIGH THROUGHPUT COMPUTING

Bring Your Own Resources (BYOR)

The members of **PATh** are working on enhancing BYOR services for commercial cloud resources

- Enable users to define, create, manage, and control usage of collections of resources purchased from commercial cloud providers. We refer to such a collection as an **Annex**
- Integrate with cloud accounts managed by the NSF CloudBank project
- Interface with workload and workflow services to manage acquisition of resources (deployment of XPoints) from commercial providers

PATh PARTNERSHIP to ADVANCE THROUGHPUT COMPUTING

Open Science Grid

HT CENTER FOR HIGH THROUGHPUT COMPUTING

Commercial Clouds offer a new level of elasticity, BYOR support must be “at scale”

PATh PARTNERSHIP to ADVANCE THROUGHPUT COMPUTING

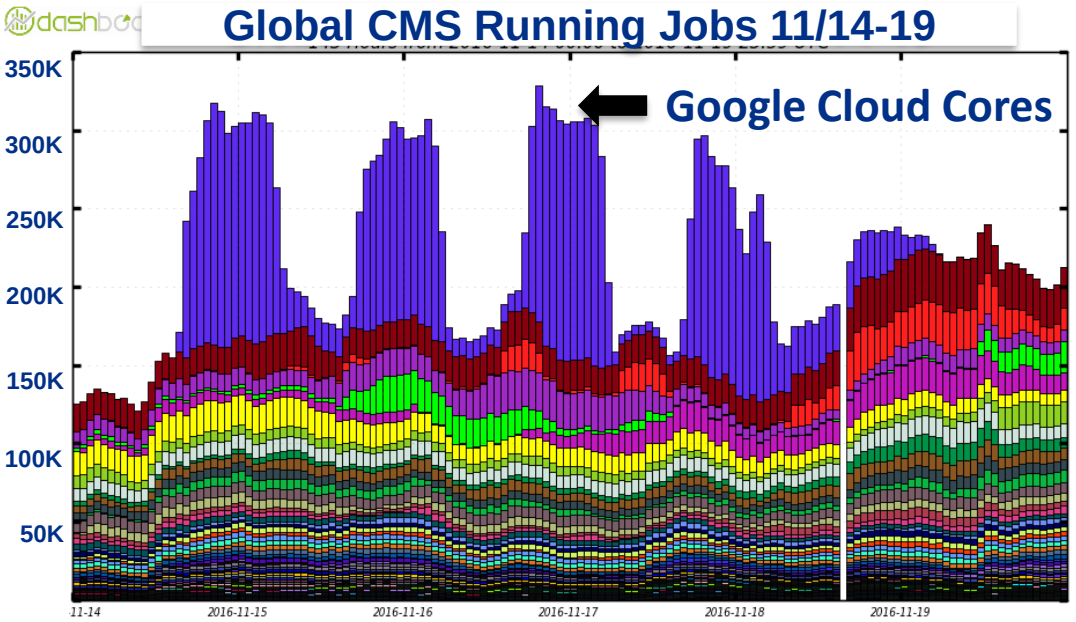
HT CENTER FOR HIGH THROUGHPUT COMPUTING

SC16: On Demand Doubling of CMS Computing Capacity

Joint project

HEPCloud (Fermilab), **HTCondor** (UW-Madison)

- **HEPCloud** provisions **Google Cloud** with **HTCondor** in two ways
  - **HTCondor** talks to Google API
  - Resources are joined into **HEP HTCondor** pool
- Demonstrated sustained large scale elasticity (>150K cores) in response to demand and external constraints
  - Ramp-up/down with opening/closing of exhibition floor
  - Tear-down when no jobs are waiting



730,172 jobs consumed 6.35M core hours produced 205M simulated events (81.8 TB) using .5PB of input data

Total cost ~$100K

Archimedes of Syracuse was a Greek mathematician, philosopher, scientist and engineer.

Give me a place to run an XPoint and I shall run your job.

Frank Würthwein is a Physics professor at UCSD and the Executive Director of the OSG

PATh PARTNERSHIP to ADVANCE THROUGHPUT COMPUTING

Open Science Grid

HT CENTER FOR HIGH THROUGHPUT COMPUTING

"Any opinions, findings, conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the Networking and Information Technology Research and Development Program."

The Networking and Information Technology Research and Development (NITRD) Program

Mailing Address: NCO/NITRD, 2415 Eisenhower Avenue, Alexandria, VA 22314

Physical Address: 490 L'Enfant Plaza SW, Suite 8001, Washington, DC 20024, USA Tel: 202-459-9674, Fax: 202-459-9673, Email: nco@nitrd.gov, Website: https://www.nitrd.gov