

SC17 Denver, Colorado

# ***Blurring the Lines: High-End Computing and Data Science***

## **Moderator**

Sandy Landsberg, DoD High Performance Computing Modernization Program (HPCMP)

## **Panelists**

Steve Conway, Hyperion Research

Satoshi Matsuoka, Tokyo Institute of Technology

Fran Berman, Rensselaer Polytechnic Institute

Michela Taufer, University of Delaware

Bob Grossman, University of Chicago

Rick Stevens, Argonne National Laboratory, University of Chicago

*Organized by the Networking and Information Technology Research and Development (NITRD) High-End Computing and Big Data Interagency Working Groups.*

# Blurring the Lines:

## High-End Computing and Data Science

**High-End Computing (HEC)** encompasses both massive computational and big data **capability** to solve computational problems of significant importance that are beyond the capability of small- to medium-scale systems. **Data science includes large-scale data analytics and visualization across multiple scales of data** from a multitude of sources. Increasingly on-demand and real-time data intensive computing, enabling real-time analysis of simulations, data-intensive experiments and streaming observations, is pushing the boundaries of computing and resulting in a convergence of traditional HEC and newer cloud computing environments. This panel will **explore challenges and opportunities at the intersection of high-end computing and data science.**

- Which markets will drive the adoption of HEC for Data Science? What new applications could arise from this convergence? What game-changers will this enable?
- What are the impacts on our current computing ecosystems and the implications for future computing ecosystems? What impact will this have on conventional workflows, architectures and new memory paradigms (supercomputers versus shared cloud computing environments), software tools and workforce development?

# Panel Logistics

- **Panelists:**

- Steve Conway, Hyperion Research
- Satoshi Matsuoka, Tokyo Institute of Technology
- Fran Berman, Rensselaer Polytechnic Institute
- Michela Taufer, University of Delaware
- Bob Grossman, University of Chicago
- Rick Stevens, Argonne National Laboratory, University of Chicago

- **Logistics:**

- Each panelists will have 7-8 minutes to present (50 minutes)
- Question and answer with audience – microphones in room and electronic on SC17 website – go to our panel webpage.





HYPERION RESEARCH

# Blurring the Lines: High-End Computing & Data Science

**SC17**

Steve Conway  
SVP-Research

# Convergence of HPC Data-Intensive Simulation and Analytics (High Performance Data Analysis)

## Modeling & Simulation

- Existing HPC users
  - Larger problem sizes
  - Higher resolution
  - Iterative methods
  - EP jobs to the cloud (Novartis)
- New commercial users
  - E.g., SMEs

## Convergence Market (2020)

**\$3.5B servers**

**\$1.6B storage**

**\$5.1B total**

## Advanced Analytics

- Existing HPC users
  - Intelligence community, FSI
  - Data-driven science/engineering (e.g., biology)
  - Knowledge discovery
  - ML/DL, cognitive, AI
- New commercial users
  - Fraud/anomaly detection
  - Business intelligence
  - Affinity marketing
  - Personalized medicine

### Drivers:

- Competition
- Complexity
- Time

Forecast:

# HPDA Market and ML/DL/AI Methods

**TABLE 1**

**Worldwide HPC AI Server Revenues vs. All HPDA Server Revenues (\$ Millions)**

	2015	2016	2017	2018	2019	2020	2021	CAGR 16-21
Total WW HPDA Server Revenues	\$1,455	\$1,845	\$2,333	\$2,830	\$3,224	\$3,488	\$4,040	17.0%
Total HPC-Based AI (DL, ML, and Other)	\$246	\$346	\$501	\$673	\$845	\$986	\$1,260	29.5%

Source: Hyperion Research 2017

# HPDA Analytics → New HPC Segments

## 1. Fraud and anomaly detection.

- Government (intelligence, cyber security)
- Industry (credit card fraud, cyber security)

## 2. Affinity Marketing.

- Discern potential customers' demographics, buying preferences and habits.

## 3. Business intelligence.

- Identify opportunities to advance market position and competitiveness

## 4. Precision Medicine

- Personalized approach to improve outcomes, control costs



# AI/Deep Learning Formative Market



*“The amount of data available today is miniscule compared to what we need for deep learning.”*

**Marti Head,  
GlaxoSmithKline**

## MARKET STATUS

- HPC has moved to the forefront of DL/AI research
- Ecosystem (including GPGPUs) formed around social media/Web giants
- DL needs massive data: not available yet in many markets
- Lack of standard benchmarks lengthens sales process
- Need for transparency ➡ HPC simulation!



# Game

**Google's DeepMind (AlphaGo) defeats the best humans.**

**“We still can't explain it...you could...review...every parameter in AlphaGo's artificial brain, but even a programmer would not glean much from these numbers because what drives a neural net to make a decision is encoded in the billions of diffuse connections between nodes.”**

*Alan Whitfield, robot ethicist, Univ. of the West of England*



# Life

**If an autonomous vehicle kills pedestrians in an accident...**

Automakers, insurance companies and auto owners will need to know why.





# **Converging HPC and BD/AI: Tokyo Tech. TSUBAME3.0 and AIST ABCI**

Satoshi Matsuoka

Professor, GSIC, Tokyo Institute of Technology /  
Director, AIST-Tokyo Tech. Big Data Open Innovation Lab /  
Fellow, Artificial Intelligence Research Center, AIST, Japan /  
Vis. Researcher, Advanced Institute for Computational  
Science, Riken

Convergence Panel

2017/11/14

Denver, Colorado, USA

# Characteristics of Big Data and AI Computing

*As BD / AI*

Graph Analytics e.g. Social Networks  
Sort, Hash, e.g. DB, log analysis  
Symbolic Processing: Traditional AI



*As HPC Task*

Integer Ops & Sparse Matrices  
Data Movement, Large Memory  
Sparse and Random Data, Low Locality



*Acceleration, Scaling*

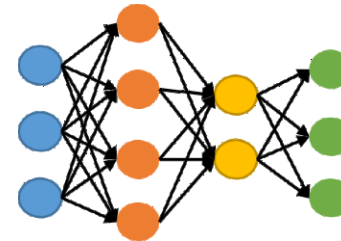
Opposite ends of HPC  
computing spectrum,  
but HPC simulation  
apps can also be  
categorized likewise



Acceleration via  
Supercomputers  
adapted to AI/BD

*As BD / AI*

Dense LA: DNN  
Inference, Training, Generation



*As HPC Task*

Dense Matrices, Reduced Precision  
Dense and well organized networks  
and Data

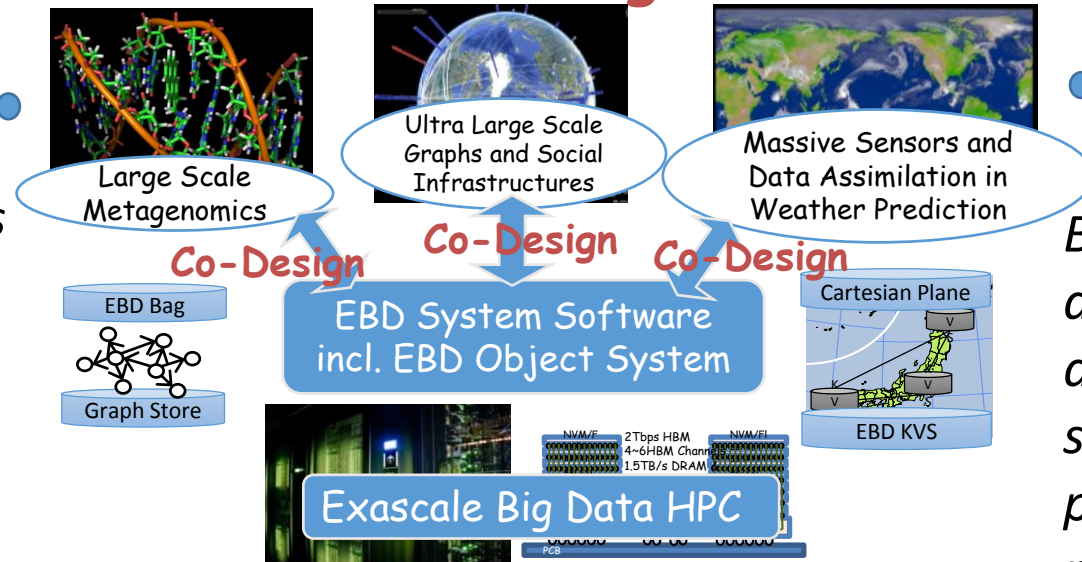


*Acceleration, Scaling*

# JST-CREST “Extreme Big Data” Project (2013-2018)

## Future Non-Silo Extreme Big Data Scientific Apps

*Given a top-class supercomputer, how fast can we accelerate next generation big data c.f. Clouds?*



*Bring HPC rigor in architectural, algorithmic, and system software performance and modeling into big data*

**Convergent Architecture (Phases 1~4)**  
**Large Capacity NVM, High-Bisection NW**

**Cloud IDC**  
Very low BW & Efficiency  
Highly available, resilient



**Supercomputers**  
Compute&Batch-Oriented  
More fragile



# EBD System Software (Matsuoka-G)

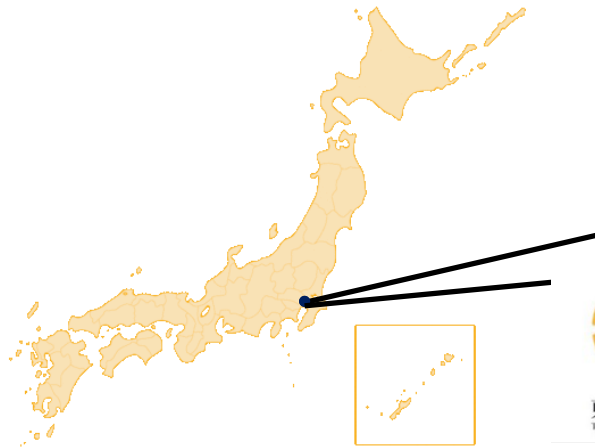
- **Big Data Algorithms for Accelerators (GPU and FPGAs, low level kernels for DNN&Graph)**
  - Fast and Memory-saving SpGEMM on GPUs
  - Accelerating SpMV on GPU by Reducing Memory Access
  - OpenCL-based High-Performance 3D Stencil Computation on FPGAs
  - Evaluating Strategies to Accelerate Applications using FPGAs
  - Accelerating Spiking Neural Networks on FPGAs
  - Directive-based Temporal-Blocking application
- **Large Scale Graph Algorithms and Sorting**
  - No.1 on Graph500 Benchmark, 5 consecutive times (collab. w/Kyushu-U, Riken etc.)
  - Distributed Large-Scale Dynamic Graph Data Store & Large-scale Graph Colouring (vertex coloring)
  - Dynamic Graph Data Structure Using Local-NVRAM
  - Incremental Graph Community Detection
  - ScaleGraph: Large-scale Graph Processing Framework w/ User-Friendly Interface
  - GPU-HykSort: Large Scale Sorting on Massive GPUs
  - XtrSort: GPU out of core sorting
  - Efficient Parallel Sorting Algorithm for Variable-Length Keys
- **Big-Data Performance Modeling and Analysis**
  - Co-locating HPC and Big Data Analytics
  - Visualizing Traffic of Large-scale Networks
  - I/O vs MPI Traffic Interference on Fat-tree Networks
  - ibprof : Low-level Profiler of MPI Network Traffic
  - Evaluation of HPC-Big Data Applications in Clouds
  - Analysis on Configurations of Burst Buffers
- **High Performance Big-Data Programming Middleware**
  - mrCUDA: Remote-to-local GPU Migration Middleware
  - Transpiler between Python and Fortran
  - Hamar (Highly Accelerated Map Reduce)
  - Out-of-core GPU-MapReduce for Large-scale Graph Processing
  - DRAGON: Extending UVM to NVMe
  - Hierarchical, User-level and ON-demand File system (HuronFS)
- **Optimizing Traffic Simulation App (Ex- Suzumura Group)**
  - Incremental Graph Community Detection
  - DeepGraph
  - Exact-Differential Traffic Simulation

## METI AIST-AIRC ABCI

as the *worlds first large-scale OPEN AI Infrastructure*

- **ABCI:** AI Bridging Cloud Infrastructure

- Top-Level SC compute & data capability for DNN (**550 AI-Petaflops**)
- Open Public & Dedicated infrastructure for AI & Big Data Algorithms, Software and Applications – **OPEN SOURCING AI DATACENTER**
- Platform to accelerate joint academic-industry R&D for AI in Japan



Univ. Tokyo Kashiwa Campus

- >550 AI-Petaflops
- < 3MW Power
- < 1.1 Avg. PUE
- Operational 2017Q4  
~2018Q1



# The “Real” ABCI – 2018Q1

- **Extreme computing power**
  - w/ 550 **AI-PFlops** (likely several 100s AI-Pflops) for AI/ML especially DNN
  - **several million speedup** over high-end PC: 1 Day training for 10,000-Year DNN training job
  - TSUBAME-KFC (1.4 AI-Pflops) x 90 users (T2 avg) min
- **Big Data and HPC converged modern design**
  - **Not just (AI-)FLOPS, but with BYTES (capacity and bandwidth)**
  - Leverage Tokyo Tech’s “TSUBAME3” design, but differences/enhancements being AI/BD centric
- **Ultra high BW & Low latency memory, network, and storage**
  - For accelerating various AI/BD workloads
  - Data-centric architecture, optimizes data movement
- **Big Data/AI and HPC SW Stack Convergence**
  - Incl. results from JST-CREST EBD
  - **Wide contributions from the PC Cluster community desirable.**
- **Ultra-Green (PUE<1.1), High Thermal (60KW) Rack**
  - Custom, warehouse-like IDC building and internal pods
  - Final “commoditization” of HPC technologies into Clouds



イメージスケッチ

# Basic Requirements for AI Cloud System

## BD/AI User Applications

Machine Learning Libraries

Graph Computing Libraries

Deep Learning Frameworks

Web Services

Python, Jupyter Notebook, R etc. + IDL

Numerical Libraries  
BLAS/Matlab

BD Algorithm  
Kernels (sort etc.)

Fortran · C · C++  
Native Codes

MPI · OpenMP/ACC · CUDA/OpenCL

Parallel Debuggers and Profilers

PFS  
Lustre · GPFS

DFS  
HDFS

RDB  
PostgreSQL

CloudDB/NoSQL  
Hbase/Mondb/Redis

SQL  
Hive/Pig

Batch Job  
Schedulers

Workflow  
Systems

Resource  
Brokers

Linux Containers · Cloud Services

Linux OS

IB · OPA  
High Capacity  
Low Latency NW

Local  
Flash+3D  
XPoint  
Storage

X86 (Xeon, Phi) +  
Accelerators e.g.  
GPU, FPGA, Lake  
Crest

## Application

- ✓ Easy use of various ML/DL/Graph frameworks from Python, Jupyter Notebook, R, etc.
- ✓ Web-based applications and services provision

## System Software

- ✓ HPC-oriented techniques for numerical libraries, BD Algorithm kernels, etc.
- ✓ Supporting long running jobs / workflow for DL
- ✓ Accelerated I/O and secure data access to large data sets
- ✓ User-customized environment based on Linux containers for easy deployment and reproducibility

## OS

## Hardware

- ✓ Modern supercomputing facilities based on commodity components



# Cutting Edge Research AI Infrastructures in Japan

## Accelerating BD/AI with HPC (w/accompanying BYTES)

(and my effort to design & build them)

**In Production**




x5.  
8

Oct. 2015  
**TSUBAME-KFC/DL** (Tokyo Tech./NEC)  
 1.4 AI-PF(Petaflops)

**In Production** x5.  
8

Mar. 2017  
**AIST AI Cloud**  
 (AIST-AIRC/NEC)  
 8.2 AI-PF




**In Production**

Aug. 2017 x11.7  
**TSUBAME3.0** (Tokyo Tech./HPE)  
 47.2 AI-PF (65.8 AI-PF w/Tsubame2.5)



**In Construction**

1H 2018 X4~6?  
**ABCI (AIST-AIRC)**  
**550 AI-PF**



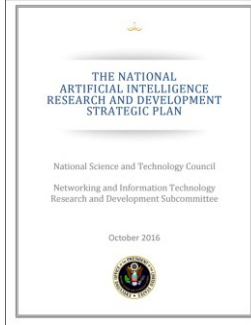
IDC under construction

1H 2019?  
**"ExaAI"**  
 ~2~3 AI-ExaFlop  
**Undergoing Engineering Study**  
 Also Post-K Multi AI-Exaflops  
 still under plans



**R&D Investments into world leading AI/BD HW & SW & Algorithms and their co-design for cutting edge Infrastructure absolutely necessary (just as is with Japan Post-K and US ECP in HPC)**

**~x400 in 3 years**  
 Built/funded



# Blurring the Lines: High-End Computing and Data Science

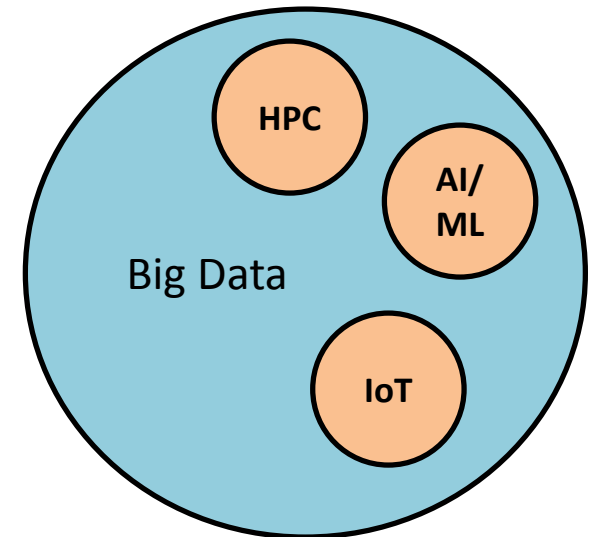
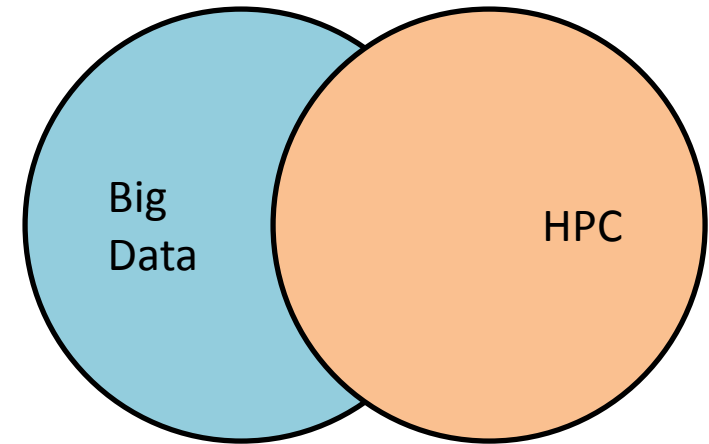
Dr. Fran Berman

Chair, Research Data Alliance / U.S.

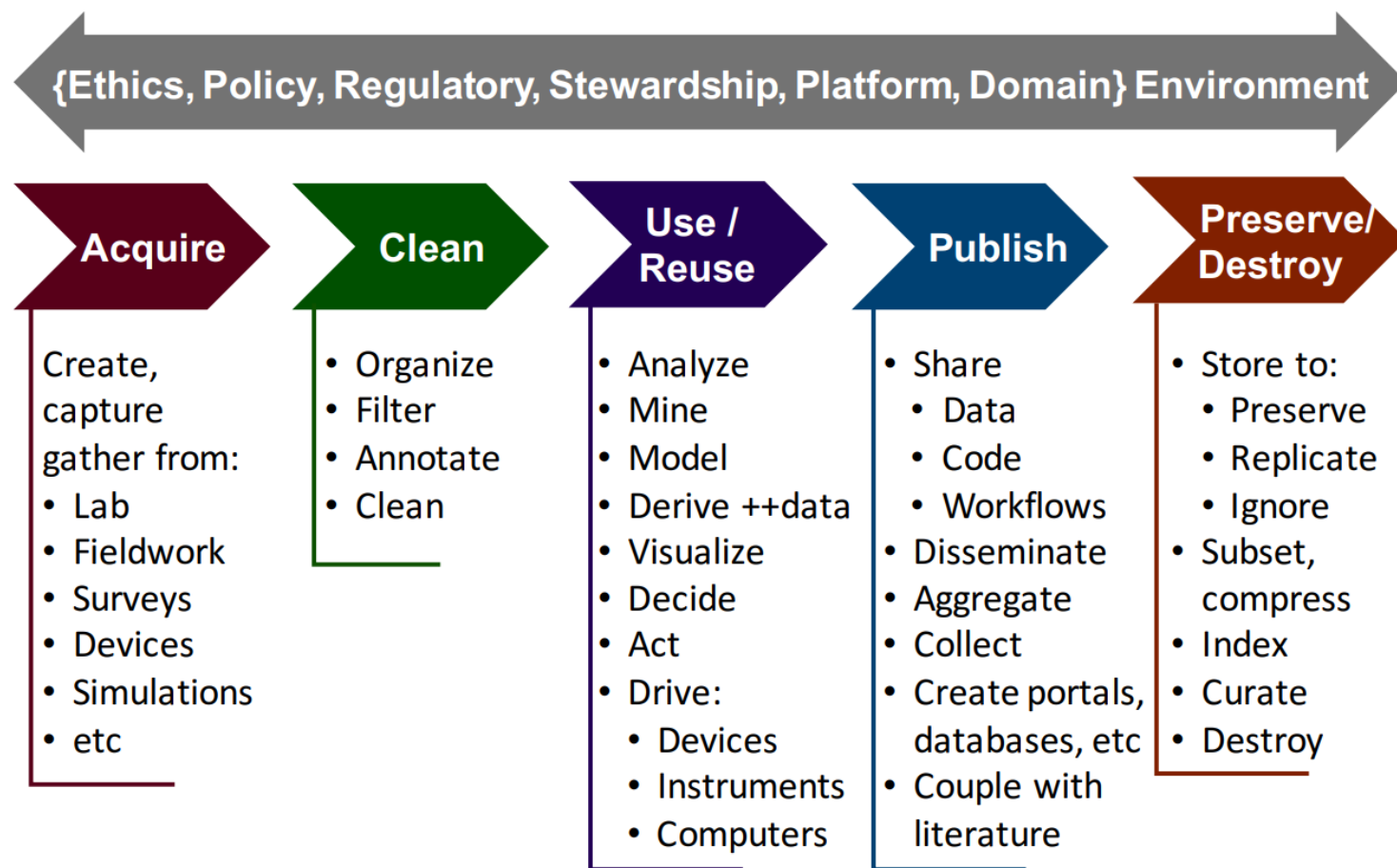
Hamilton Distinguished Professor of CS, RPI

# Thinking Big about Data

- **Increasing expansion of data science:**
  - Data expanding functionality and increases the potential for innovation in the areas it is associated with.
  - Data science seen as cross-cutting area with impact in virtually every domain and sector.
  - *Big Data* broadly interpreted.
- **Goal of Big Data efforts is big insights.**
  - From a data perspective, HPC is one of many technologies needed to drive Big Data innovation.



# Data community focused on broad set of themes in the Data Life Cycle



# Data Science Development

## Key areas for data science expansion of interest to NSF

- **Data science training and curriculum**

- Where on campus does data science “live”?
- How can we train new data scientists and data-savvy professionals?

- **Data science research**

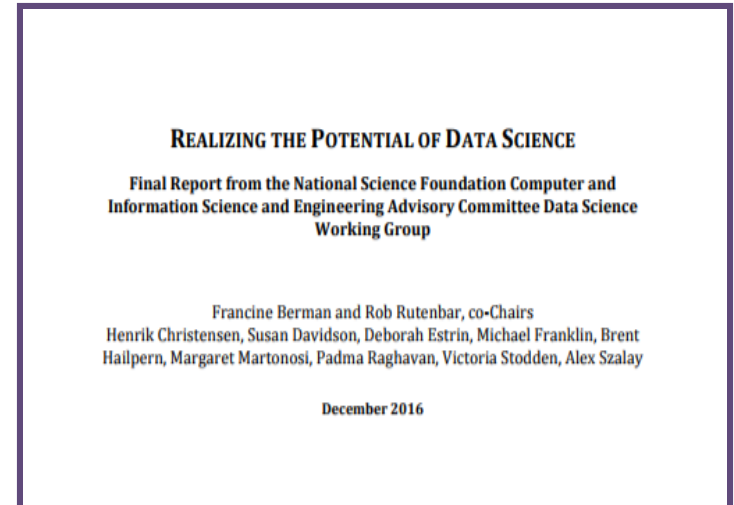
- How can we better use data to gain insights?
- How do we make data systems more robust, capable, secure?
- What policy, ethics, practice needed to get the most from data?

- **Data science infrastructure**

- How do we strengthen organizational and institutional infrastructure to support data science and data analysis?
- What stewardship, preservation, and tools infrastructure is needed to ensure data use, re-use and reproducibility?

- **Data Futures**

- How to encourage innovation for new data-driven areas?



<https://www.nsf.gov/cise/ac-data-science-report/CISEACDataScienceReport1.19.17.pdf>

# Data Futures: Internet of Things -- New applications focusing on enhancing people through technology, and technology through intelligence



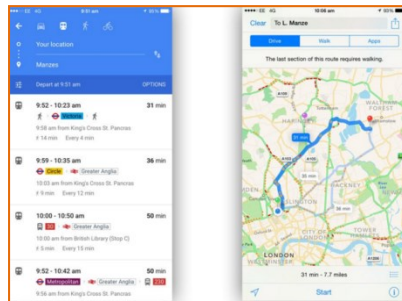
Adaptive Systems



Customization / Personalization



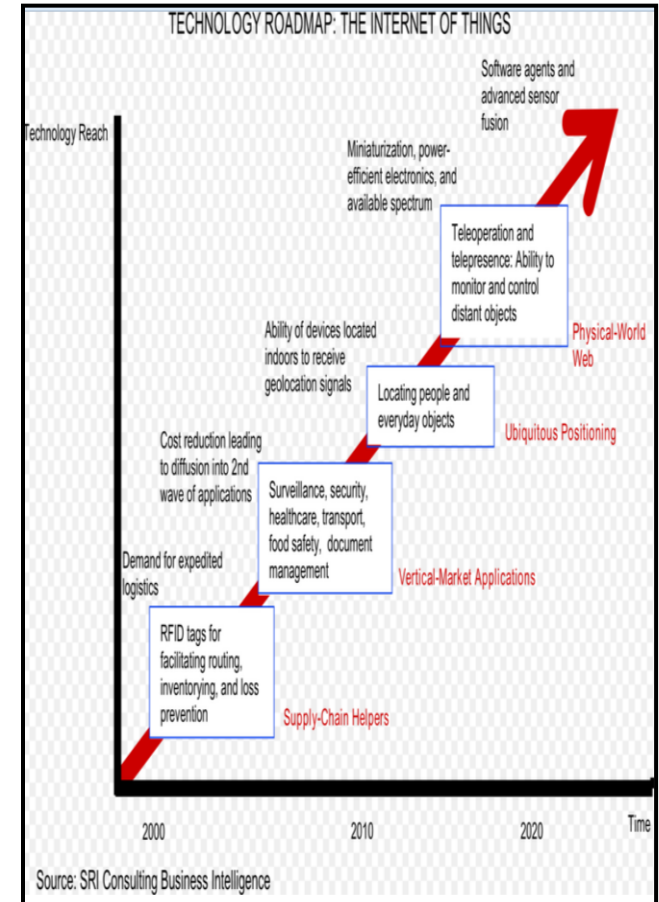
Smart Technologies



Optimization



Monitoring



# Blurring the Lines

- **Data goal (insight) vs. HPC goal (scale)**
  - Lines blurred when scale is needed for insight [private sector]
  - Lines blurred when data a stakeholder priority [academics]
  - Lines blurred when the problem best solved with data volume and at scale (e.g. earthquake simulation) [users]
  - Lines blurred when tools, infrastructure, technologies relevant to a broader set of environments, problems, users
- **Optimizing for innovation:**
  - What are the goals?
  - Who are the beneficiaries?
  - What are the metrics of success?
- **Backwards engineering from problems blurs technology silos**
- **Backwards engineering from leadership expectations strengthens silos**







# ***Challenges in Big Data Computing on HPC Platforms***

Michela Taufer

Computer and Information Sciences

University of Delaware

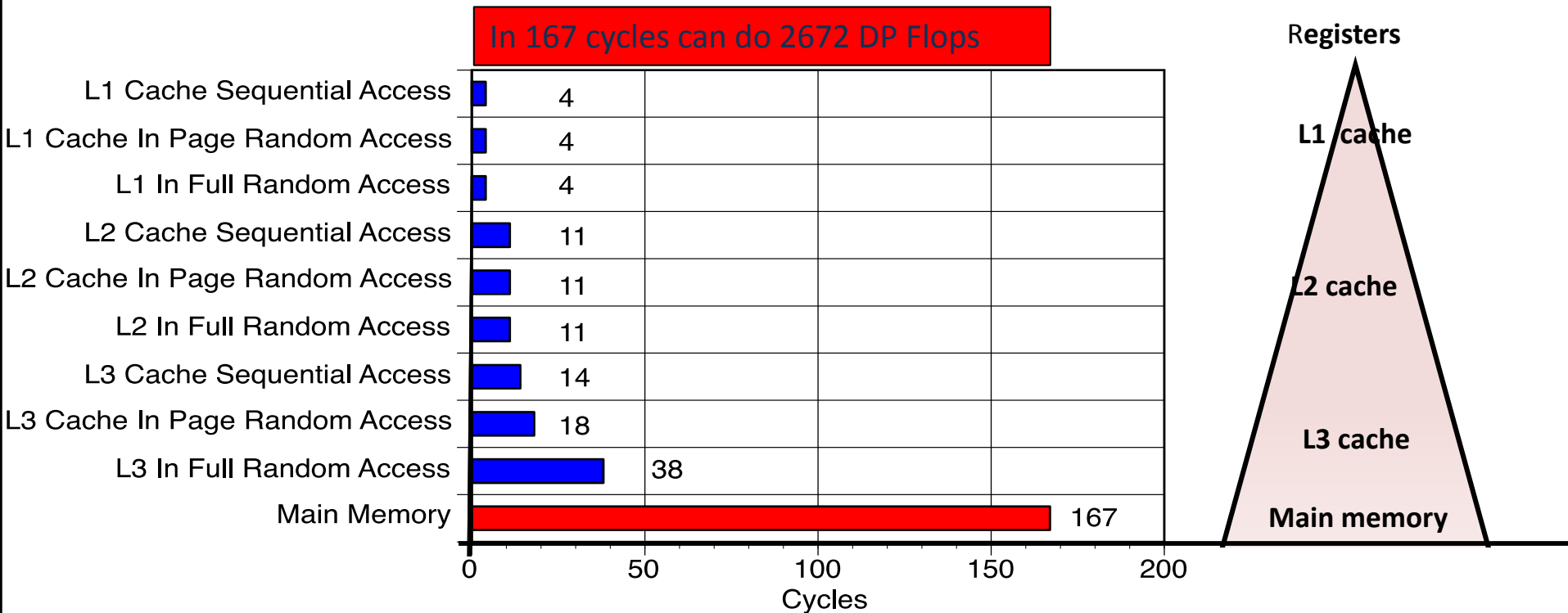
Newark, Delaware, USA





## The Cost of Data Movement

- Today's floating point operations are inexpensive



- Data movement is very expensive

*Courtesy of Jack Dongarra, UTK and ORNL, 2017*



## Perspectives

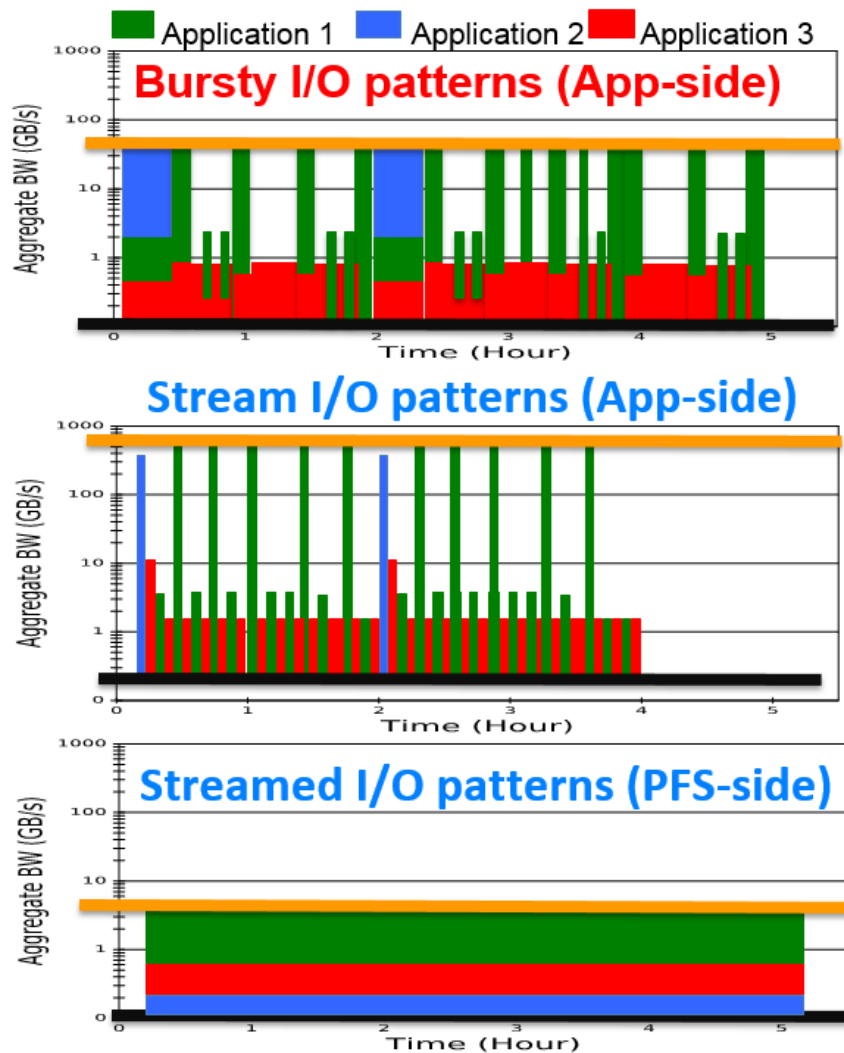
The scientist:

*“Storage technologies are advancing [...] and it is really not clear at all [to me] that especially distributed storage platforms would not be able to handle [...] petabyte data sets”*

*Anonymous Feedback*

The computer architect:

*“[...] there will be burst buffers on the DOE machines which will give applications much faster I/O [...]”*

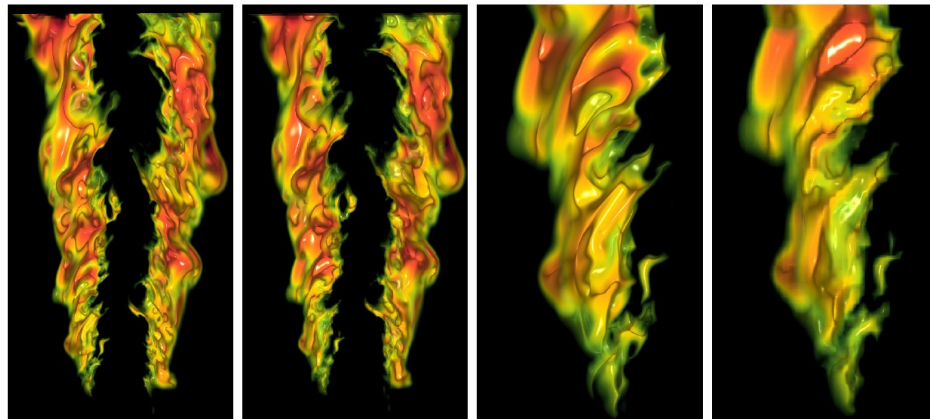
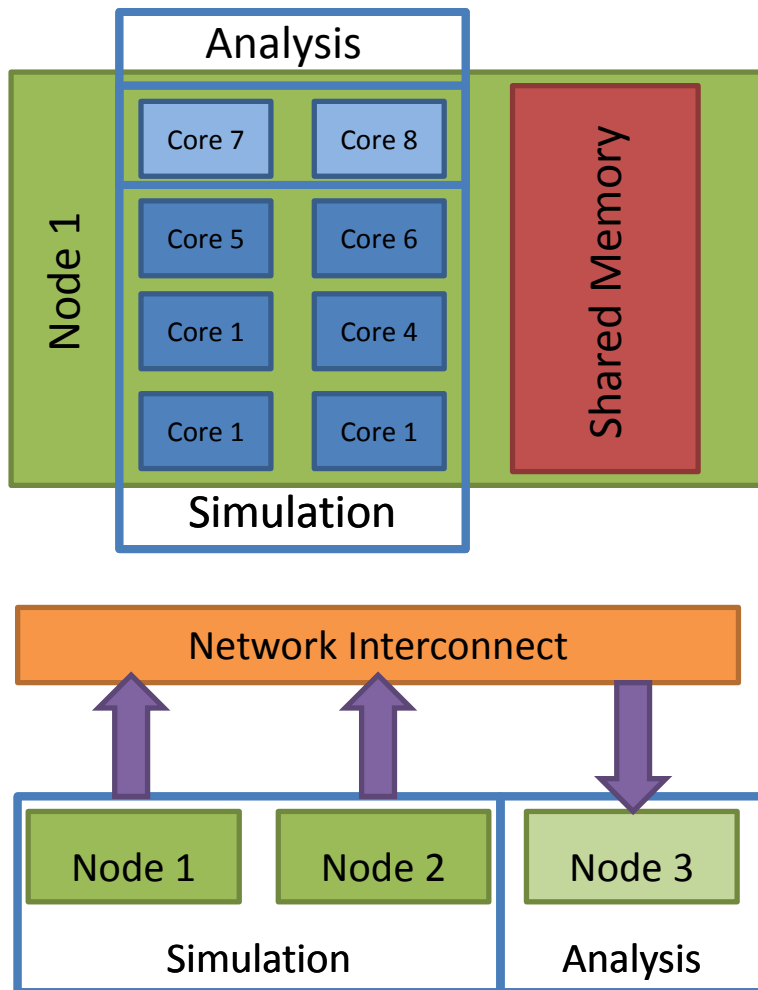


## The Burst Buffer “Revolution”

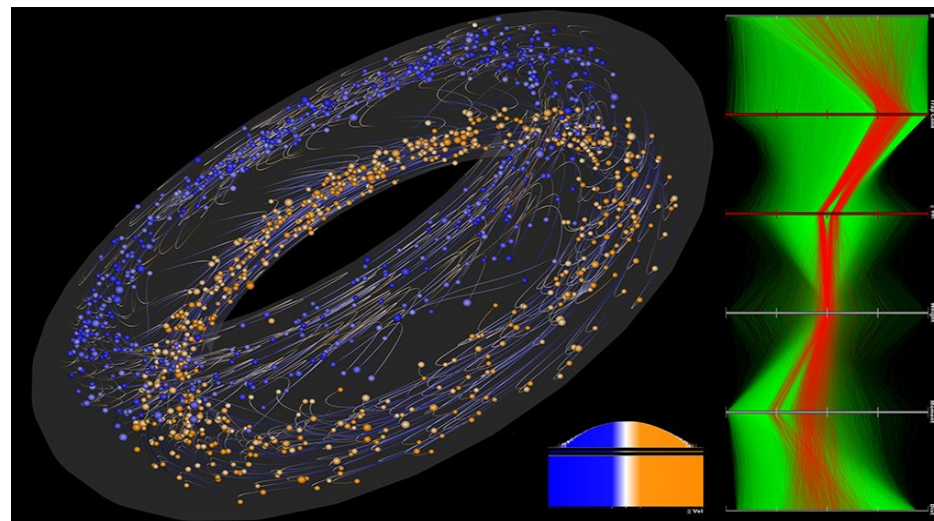
- Burst Buffers are not the magic I/O silver bullet
  - I/O contention still a problem if we exceed the BB capability
  - BBs do **NOT** help uploading data from storage for analysis and visualization
- The next “true” revolutions:
  - Algorithms for *in situ* and *in transit* analytics including ML
  - Workflows for compute and data integration



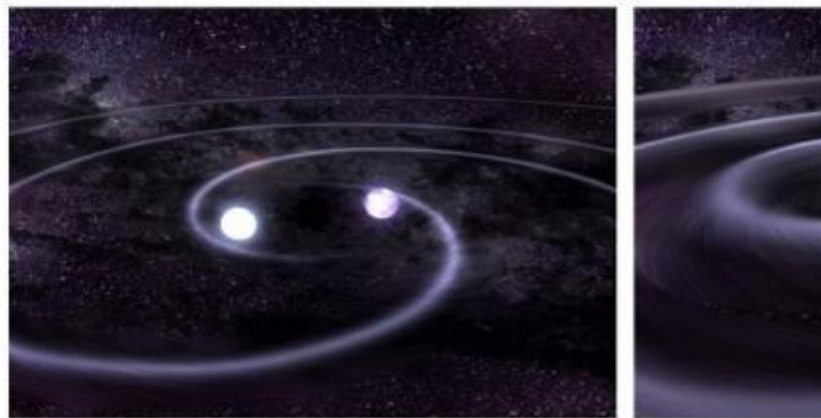
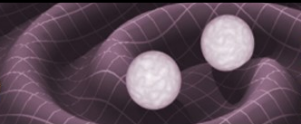
## *In-situ and In-transit Analysis*



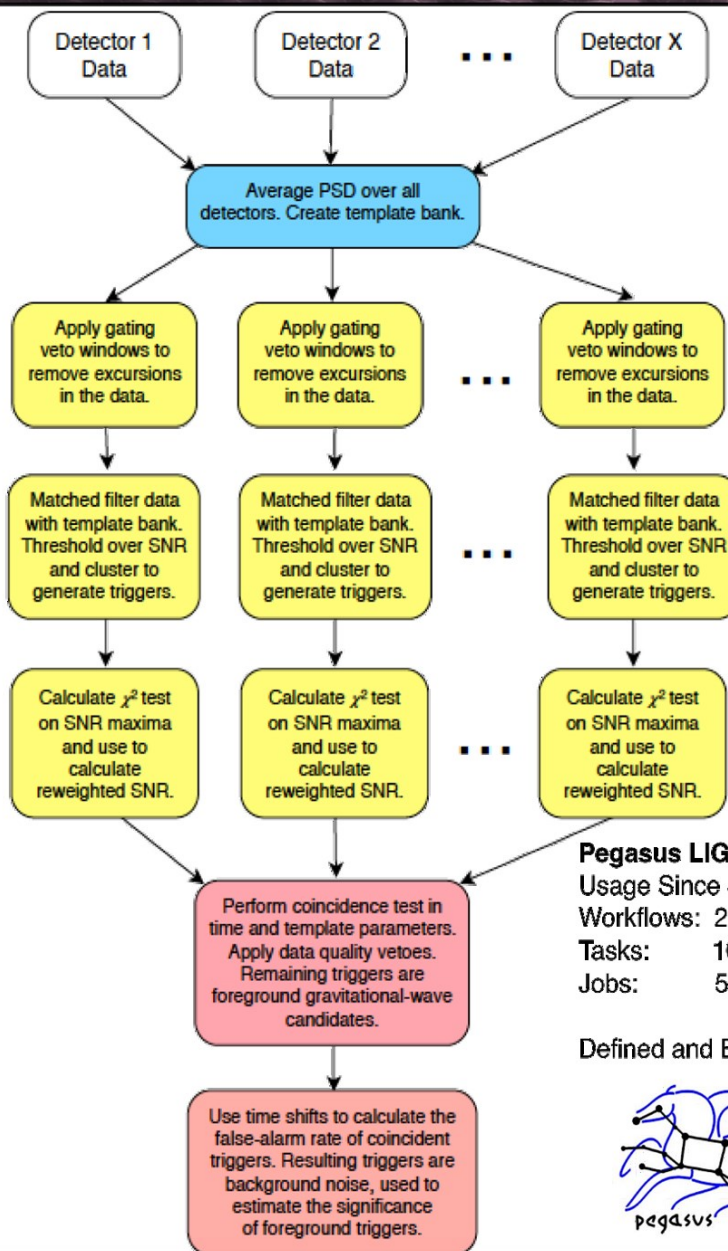
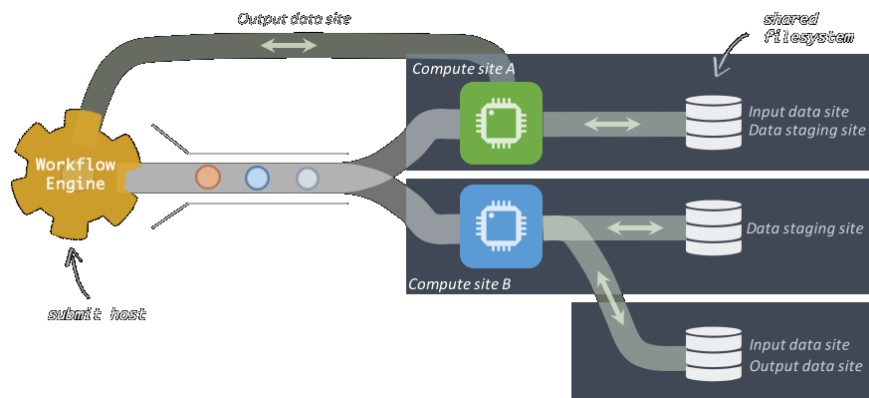
Bennett, Janine C., et al. "Combining in-situ and in-transit processing to enable extreme-scale scientific analysis." *High Performance Computing, Networking, Storage and Analysis (SC), 2012 International Conference for*. IEEE, 2012.



Abbasi, Hasan, et al. "Datastager: scalable data staging services for petascale applications." *Cluster Computing* 13.3 (2010): 277-290.



*"The inspiral and merger of two neutron stars, as illustrated here, should produce a very specific gravitational wave signal, but the moment of the merger should also produce electromagnetic radiation that's unique and identifiable as such.", credit LIGO*



First  
detection  
workflow  
statistics

**Pegasus LIGO PyCBC Workflow**  
Usage Since Sept 2015  
Workflows: 20,942  
Tasks: 107,576,294  
Jobs: 55,915,928

Defined and Executed by Pegasus



# Workflows for Compute

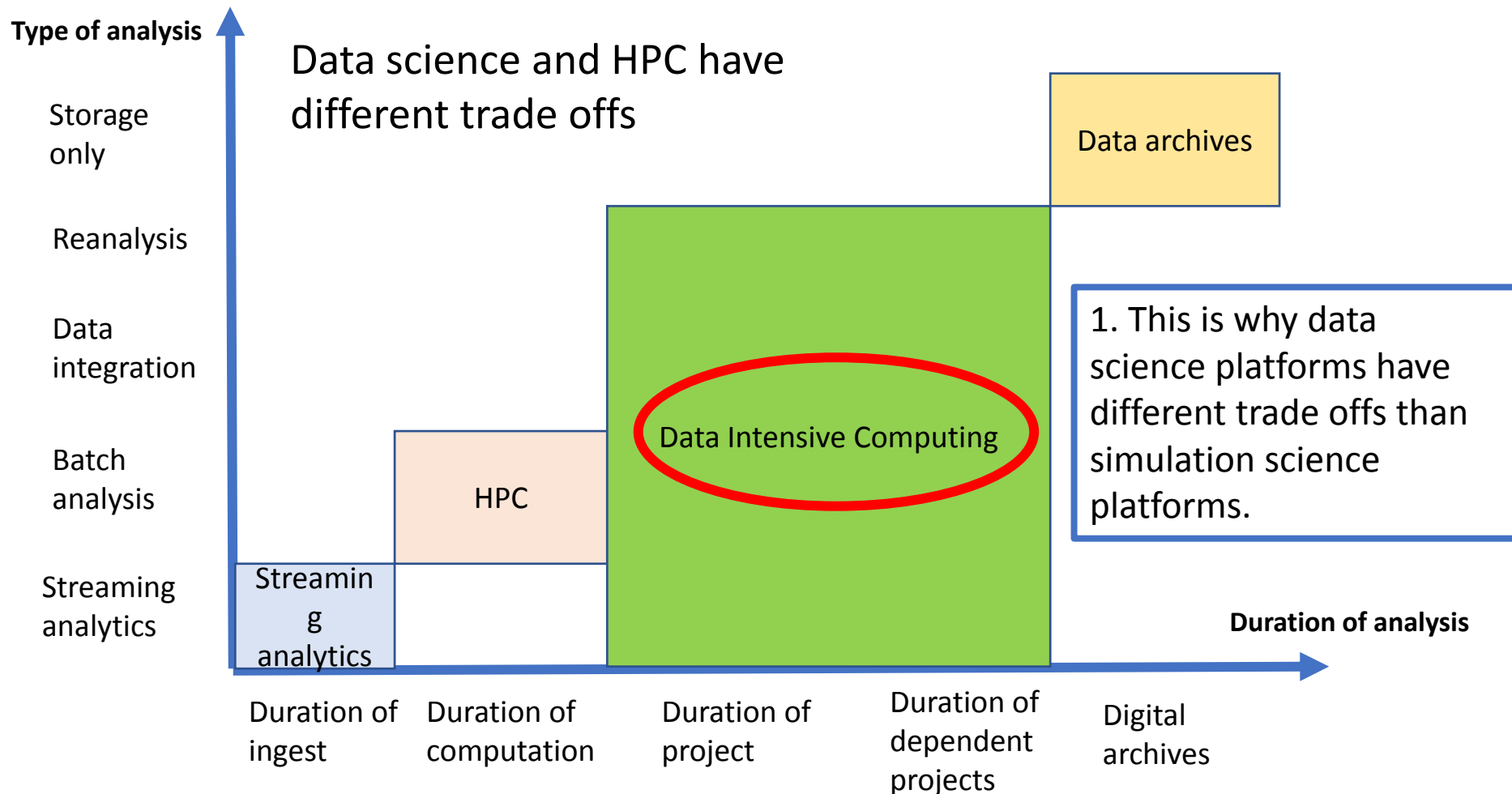


# Blurring the Lines High-End Computing & Data Science: The Data Commons Perspective

Robert L. Grossman  
University of Chicago  
& Open Commons Consortium

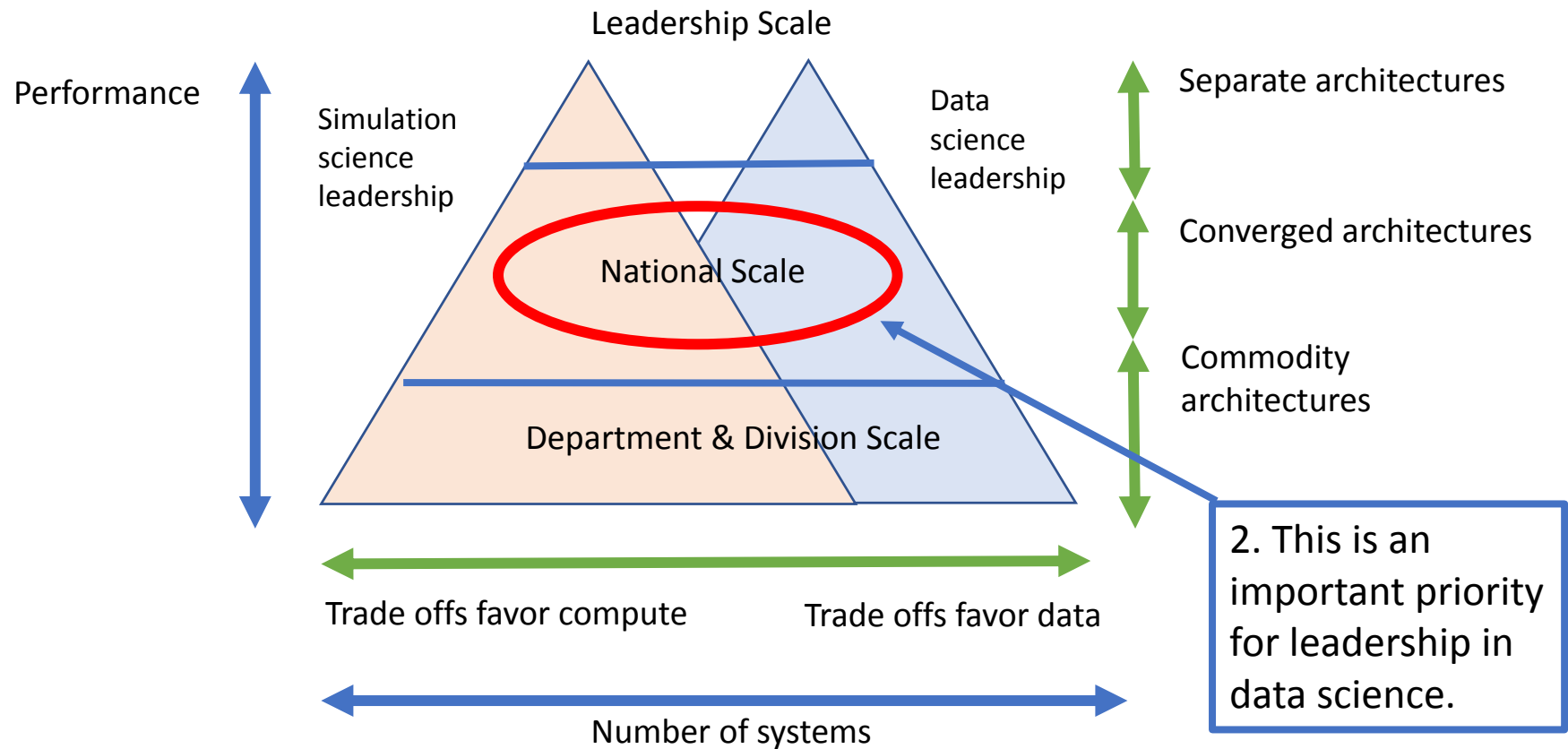
SC 17  
November 15, 2017





Adapted from: Future Directions for NSF Advanced Computing Infrastructure to Support U.S. Science and Engineering in 2017-2020, National Academy Press, DOI: 10.17226/21886, 2016

# Two Branscomb Pyramids





# Data Commons

3. Think of large scale data commons as **national scale platforms** for data science.

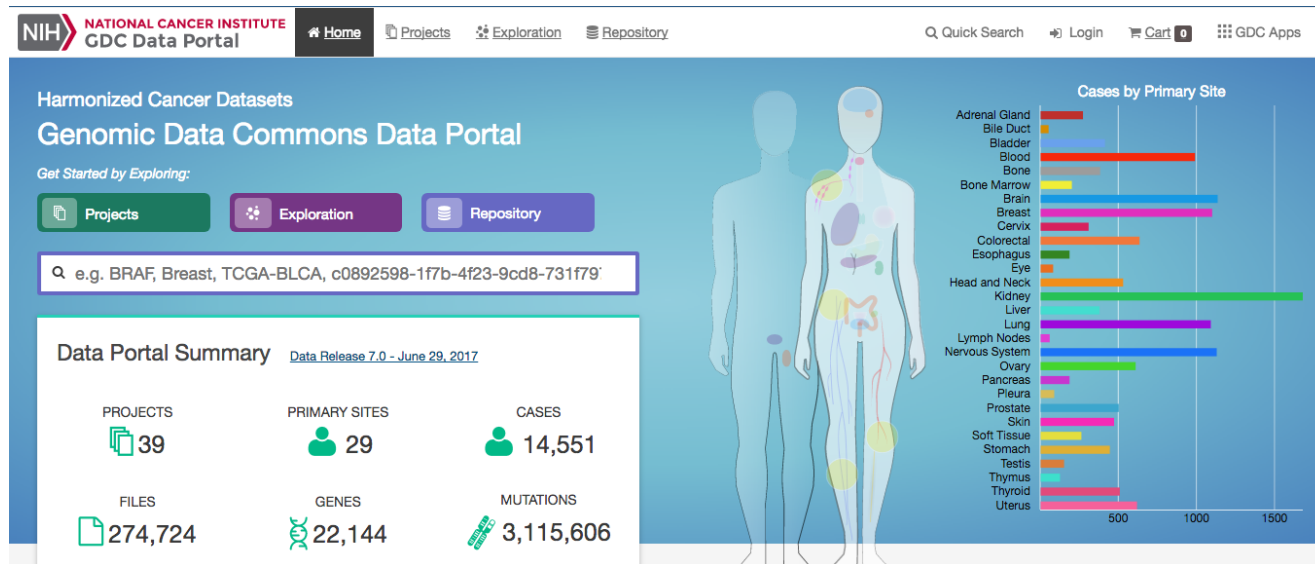


Data commons are systems that manage, analyze and share the data in a discipline or field.

Data commons co-locate data, storage and computing infrastructure with commonly used services, tools & apps for analyzing and **sharing data** to create an **interoperable** resource for the research community.\*

\*Robert L. Grossman, Allison Heath, Mark Murphy, Maria Patterson and Walt Wells, A Case for Data Commons Towards Data Science as a Service, IEEE Computing in Science and Engineer, 2016. Image: a Google data center from: [www.google.com/about/datacenters/](http://www.google.com/about/datacenters/).

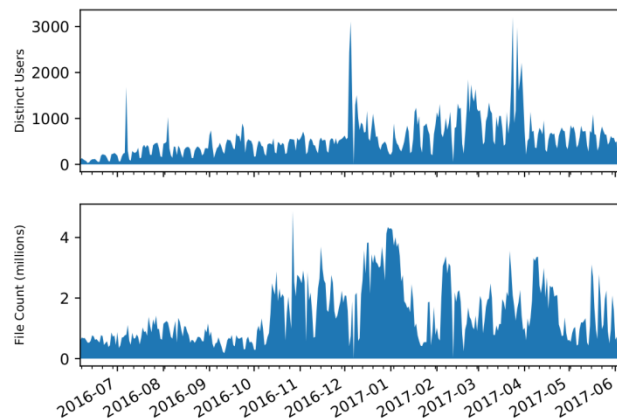
# NCI Genomic Data Commons\*



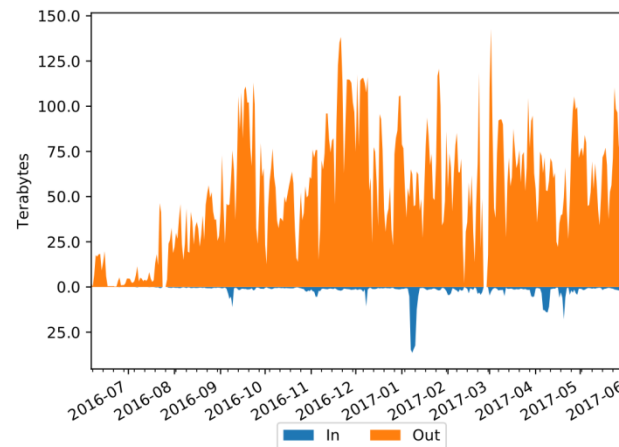
- Launched in 2016 with over 4 PB of data. Over 10 PB today.
- Used by 1500 - 2000+ users per day.
- Based upon an open source software stack that can be used to build other data commons.

\*See: NCI Genomic Data Commons: Grossman, Robert L., et al. "Toward a shared vision for cancer genomic data." New England Journal of Medicine 375.12 (2016): 1109-1112.

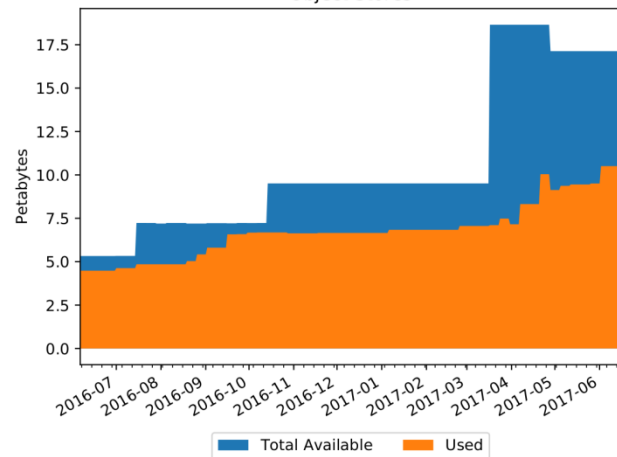
GDC User Daily Stats



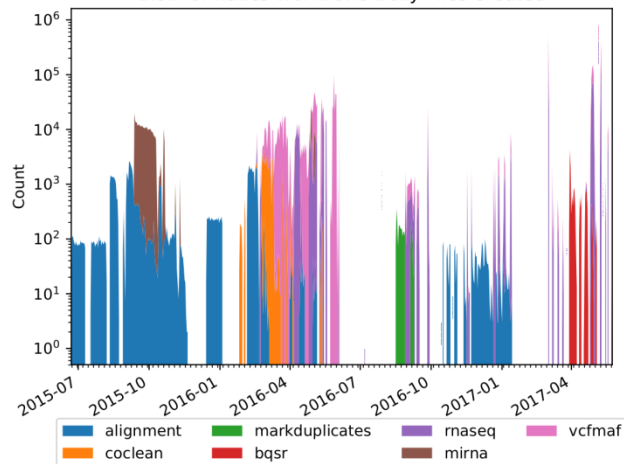
Daily Network Traffic



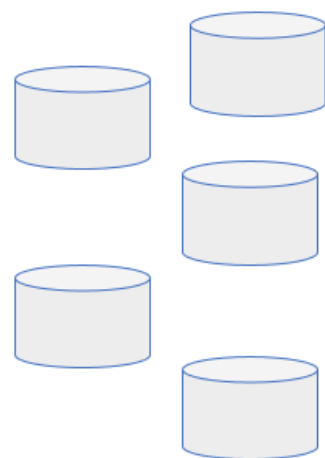
Object Stores



Bioinformatics Workflows Daily Files Created



4. HPC platforms are “destinations.”  
Data science platforms are interop resources.



**Object-based storage with access control lists**



Scalable light weight workflow



Database services

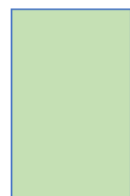


Community data products



Portals for accessing & submitting data

**Data Commons 1**



**Data Commons 2**



**APIs**



Apps



Notebooks



Apps

**Apps & Notebooks**



Workspaces



Workspaces

**Workspaces**

Data Commons Framework Services (Digital ID, Metadata, Authentication, Auth., etc.) that support multiple data commons.

**Data Commons Framework Services**

5. Both data clouds and data commons will benefit from HEC, especially as it moves to the data center.

## Databases

1982 - present



- Data repository
- Researchers download data.

## Platforms for data science

### Data Clouds

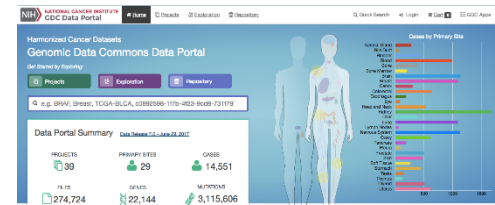
2010 - 2020



- Supports large data with **cloud computing**
- Researchers can analyze data with collaborative tools (**workspaces**) – i. e. data does **not** have to be downloaded)

### Data Commons

2014 - 2024



- Supports large data
- Workspaces
- **Common data models**
- **Core data services**
- **Harmonized data**
- **Governance**

# “Exascale: Simulation, Data and Learning”

Rick Stevens

Argonne National Laboratory

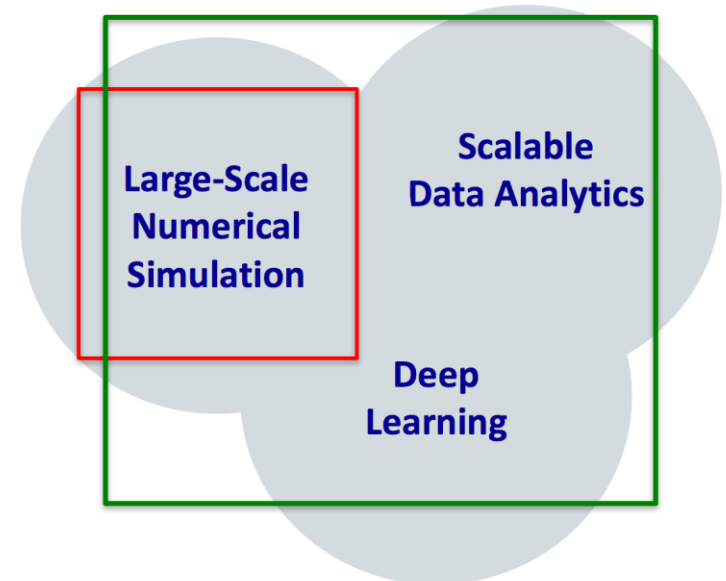
The University of Chicago



Crescat scientia; vita excolatur

# Big Picture

- Mix of applications is changing
- must support  $\Rightarrow$  Simulation, Data Analytics, and Machine Learning “AI”
- Many projects are combining all three modalities
  - Cosmology
  - Cancer
  - Materials Design
  - Climate
  - Drug Design



# Aurora 21

- Argonne's Exascale System
- Balanced architecture to support three pillars
  - Large-scale Simulation (PDEs, traditional HPC)
  - Data Intensive Applications (science pipelines)
  - Deep Learning and Emerging Science AI
- Enable integration and embedding of pillars
- Integrated computing, acceleration, storage
- Towards a common software stack



# Argonne Targets for Exascale

## Simulation Applications

- Materials Science
- Cosmology
- Molecular Dynamics
- Nuclear Reactor Modeling
- Combustion
- Quantum Computer Simulation
- Climate Modeling
- Power Grid
- Discrete Event Simulation
- Fusion Reactor Simulation
- Brain Simulation
- Transportation Networks

## Big Data Applications

- APS Data Analysis
- HEP Data Analysis
- LSST Data Analysis
- SKA Data Analysis
- Metagenome Analysis
- Battery Design Search
- Graph Analysis
- Virtual Compound Library
- Neuroscience Data Analysis
- Genome Pipelines

## Deep Learning Applications

- Drug Response Prediction
- Scientific Image Classification
- Scientific Text Understanding
- Materials Property Design
- Gravitational Lens Detection
- Feature Detection in 3D
- Street Scene Analysis
- Organism Design
- State Space Prediction
- Persistent Learning
- Hyperspectral Patterns

# Differing Requirements?

## Simulation Applications

- 64bit floating point
- Memory Bandwidth
- Random Access to Memory
- Sparse Matrices
- Distributed Memory jobs
- Synchronous I/O multinode
- Scalability Limited Comm
- Low Latency High Bandwidth
- Large Coherency Domains help sometimes
- O typically greater than I
- O rarely read
- Output is data

## Big Data Applications

- 64 bit and Integer important
- Data analysis Pipelines
- DB including No SQL
- MapReduce/SPARK
- Millions of jobs
- I/O bandwidth limited
- Data management limited
- Many task parallelism
- Large-data in and Large-data out
- I and O both important
- O is read and used
- Output is data

## Deep Learning Applications

- Lower Precision (fp32, fp16)
- FMAC @ 32 and 16 okay
- Inferencing can be 8 bit
- Scaled integer possible
- Training dominates dev
- Inference dominates pro
- Reuse of training data
- Data pipelines needed
- Dense FP typical SGEMM
- Small DFT, CNN
- Ensembles and Search
- Single Models Smallish
- I more important than O
- Output is models

# Aurora 21 Exascale Software

- Single Unified stack with resource allocation and scheduling across all pillars and ability for frameworks and libraries to seamlessly compose
- Minimize data movement: keep permanent data in the machine via distributed persistent memory while maintaining availability requirements
- Support standard file I/O and path to memory coupling for Sim, Data and Learning
- Isolation and reliability for multi-tenancy and combining workflows

# Panelists

- Steve Conway, Hyperion Research
- Satoshi Matsuoka, Tokyo Institute of Technology
- Fran Berman, Rensselaer Polytechnic Institute
- Michela Taufer, University of Delaware
- Bob Grossman, University of Chicago
- Rick Stevens, Argonne National Laboratory, University of Chicago

*"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](mailto:nco@nitrd.gov), Website: <https://www.nitrd.gov>

