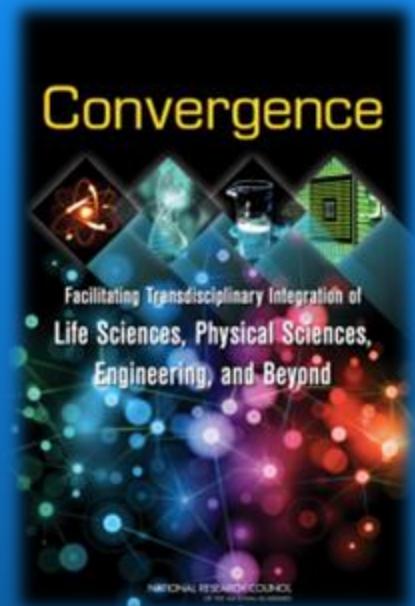
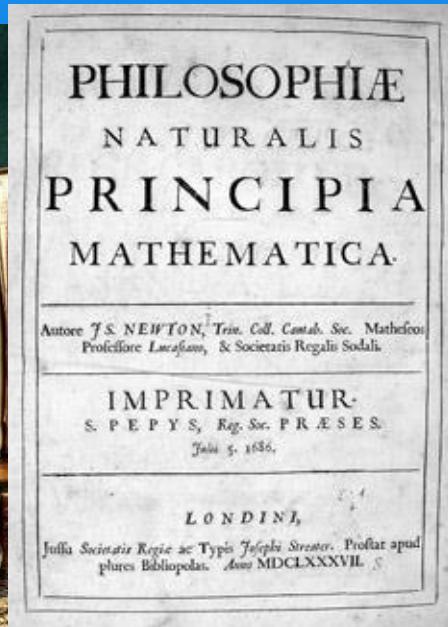
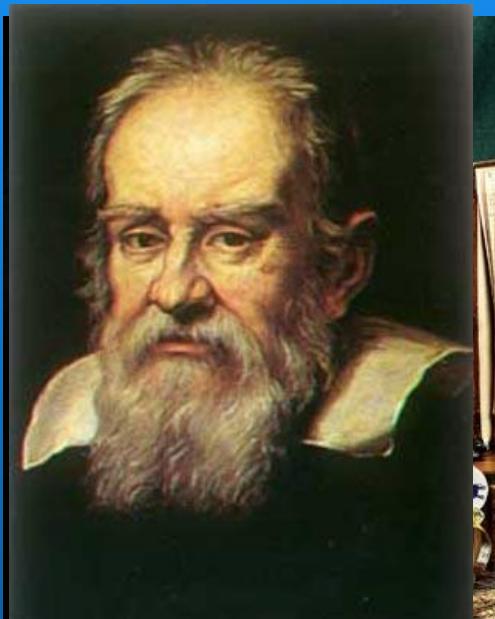


Trends in Computational and Data-enabled Science

Edward Seidel

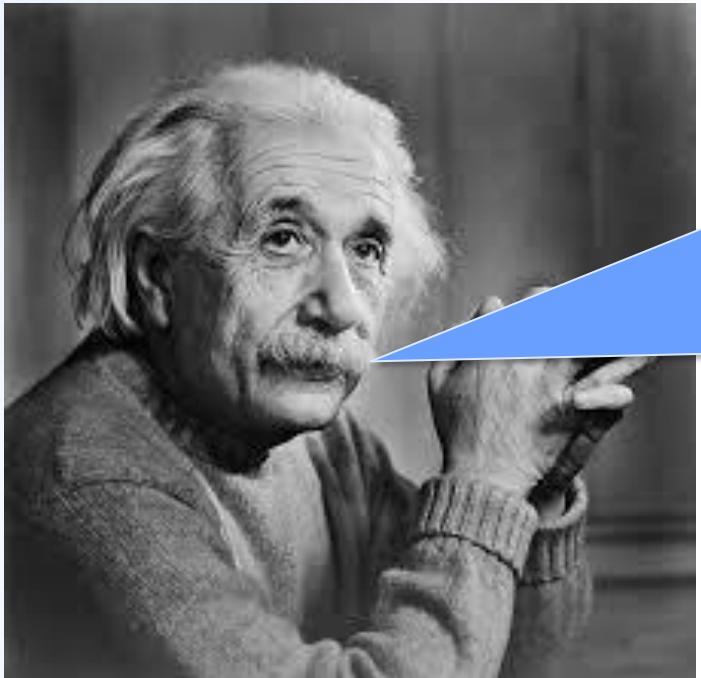
President, University of Wyoming

(Former Director, National Center for Supercomputing Applications)





*Collaborations Grow as Problems Become
More Complex; Systems and Data Sizes, too!*
TRENDS IN SCIENCE

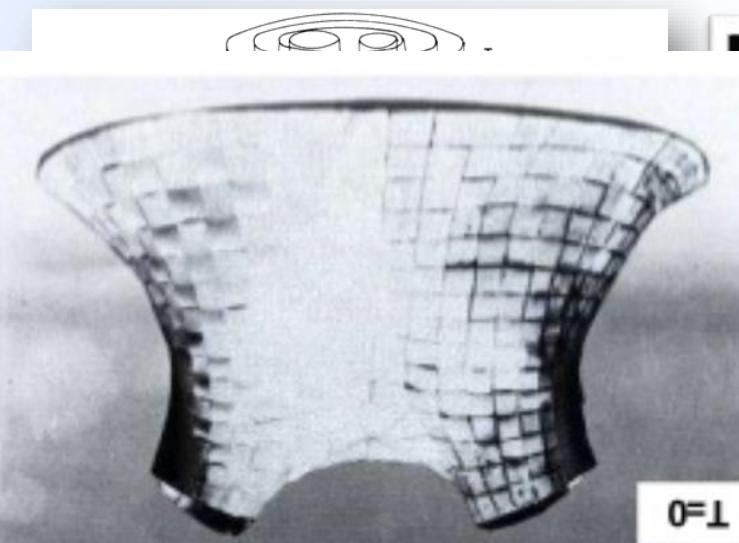


Three tech trends
made this possible:

- Instrumentation
 - HPC
 - Data Science
- They are intricately
connected!
- AI transforms it all!

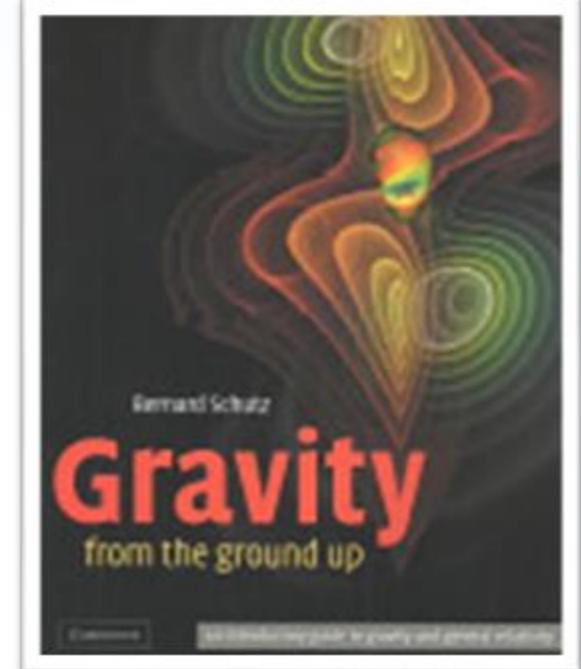
Precisely 100 years ago, Einstein predicted gravitational waves. Feb 11, 2016 their historic discovery was announced!

2 Black Hole Collision: Changing Culture of Science



Smarr, 8th Texas Symposium, p. 597 (1977)
Eppley, Ph.D. Thesis (1975), p.239

Figure 17. The collision of two Black Holes. The event horizons ∂B_1 and ∂B_2 merge to form the event horizon ∂B_3 . The apparent horizons ∂T_2 do not merge but are enveloped by a new apparent horizon ∂T_3 .



1600-1972:
Galileo - Hawking.
1 person, no
computer, 50 KB

1994: 10 people,
NCSA Cray Y-MP,
50MB



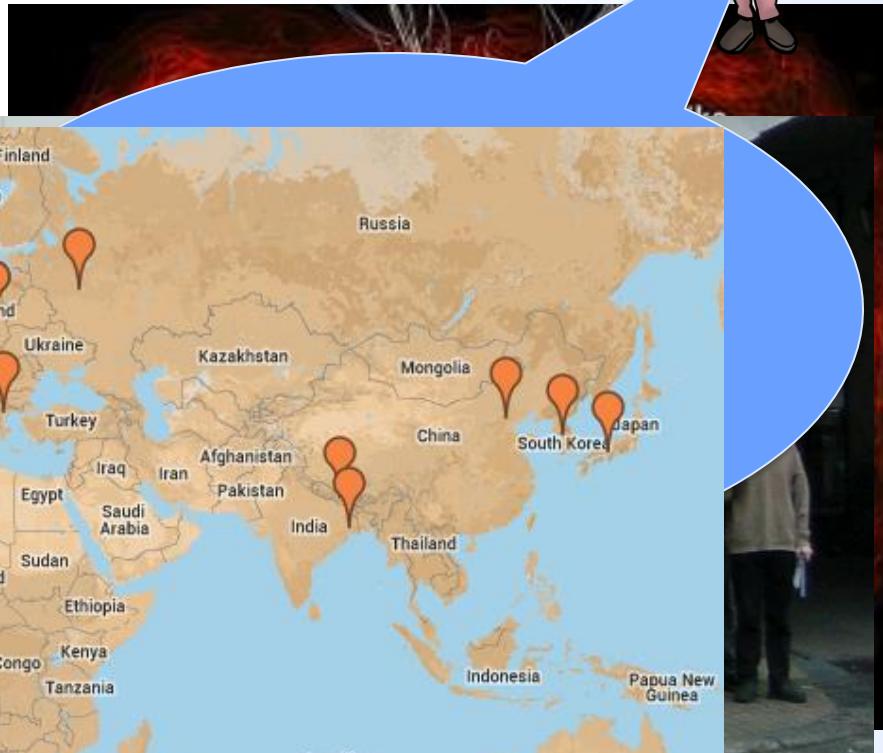
1998: 15 people,
NCSA Origin,
50GB



Community Einstein Toolkit



“Einstein Toolkit : open software for astrophysics to enable new science,

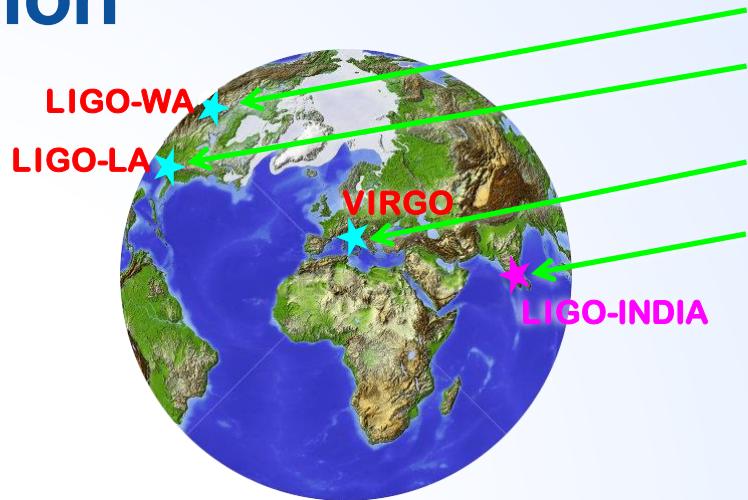


Gro
with
ific
an



LIGO Scientific Collaboration

- 1004 Authors on the breakthrough paper...
- How do they collaborate?
- They have a data sharing agreement and services to support them



Gravitational Wave Astrophysics

Obs

The New York Times

LAST UPDATE: 1:19 PM, THURSDAY, FEBRUARY 11, 2016

Stakes in Nevada Rise for Hillary Clinton as Caucuses Near

By ADAM NAGOURNEY

After his defeat of Mrs. Clinton on Tuesday, Bernie Sanders will have a chance to prove himself in a state as racially diverse as Iowa and New Hampshire are not.

Bernie Sanders Intrigues a South Carolina Town That Loves Hillary Clinton

By RICHARD FAUSSET

In Orangeburg, who said it would under Clinton are look at her



Sign in

News Sport Weather Shop Earth Travel More



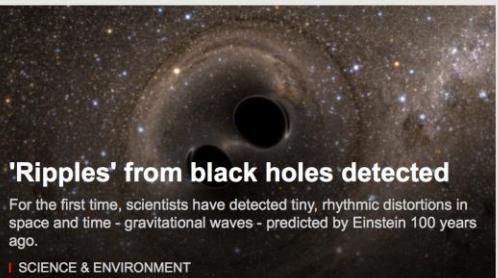
Video | LIGO Hears a Ripple in Space-Time About a hundred years ago, Einstein predicted the existence of gravitational waves, but until now, they were undetectable.

OUT THERE

Scientists Detect Gravitational Waves, Proving Einstein Right

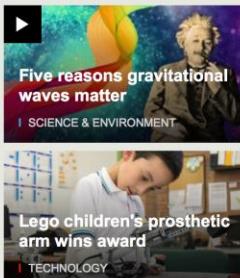
10:30 AM ET

Scientists say they heard the faint chirp of two black holes colliding a billion light-years away, fulfilling Einstein's general theory of relativity.



For the first time, scientists have detected tiny, rhythmic distortions in space and time - gravitational waves - predicted by Einstein 100 years ago.

SCIENCE & ENVIRONMENT



Five reasons gravitational waves matter
SCIENCE & ENVIRONMENT



Lego children's prosthetic arm wins award
TECHNOLOGY



Then thrive in it.

The News-Gazette

THURSDAY, FEBRUARY 11, 2016 | TODAY'S PAPER

NEWS
SHOP

SPORTS
WEEKLY ADS

BUSINESS

LIVING

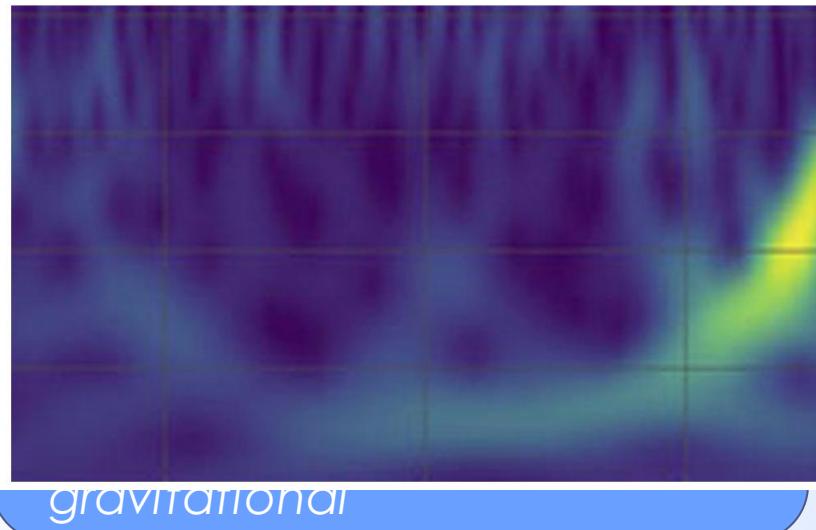
A&E

BLOGS

CLASSIFIED

Discovery confirms Einstein prediction; NCSA played role

"We can gladly say that Einstein is right, and that the beautiful mathematical framework he developed to describe gravity is valid even in the most extreme environments."



gravitational

But HPC is only part of greater ecosystem!

**BREAKTHROUGH HPC SYSTEMS
SUPPORT BREAKTHROUGH
SCIENCE**

Cray XE6/XK7 - 276 Cabinets

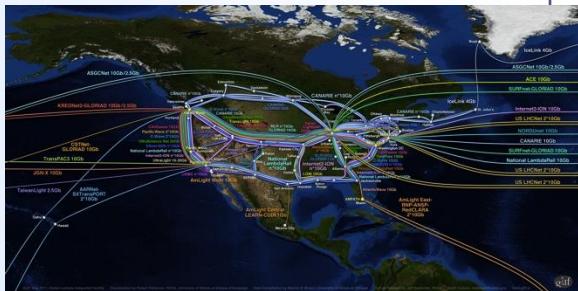
XE6 Compute Nodes - 5,688 Blades – 22,640 Nodes –
362,240 FP (bulldozer) Cores – 724,480 Integer Cores
4 GB per FP core

XK7 GPU Nodes
1,056 Blades – 4,224 Nodes
33,792 FP Cores - 11,354,112 cuda cores
– 4,224 K20X GPUs, 4 GB per FP core

Blue Waters Computing System (2012-2021)



Aggregate Memory – 1.6 PB 13.2 Peak PF



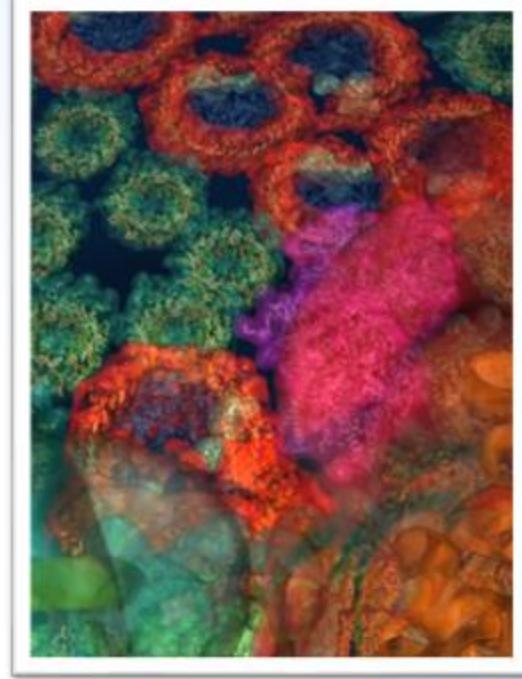
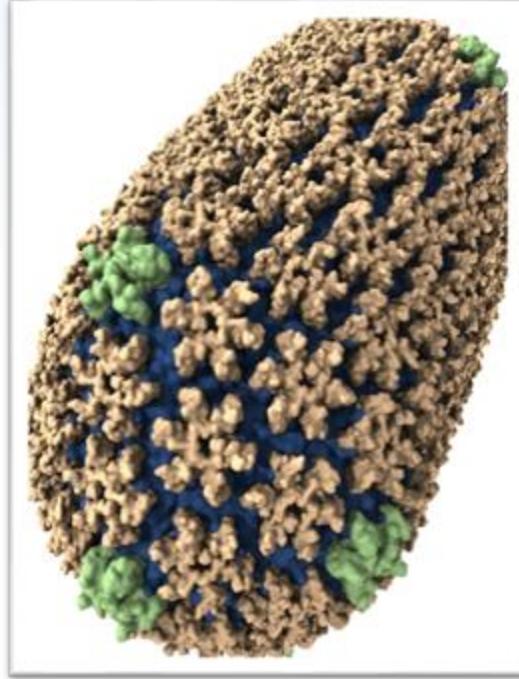
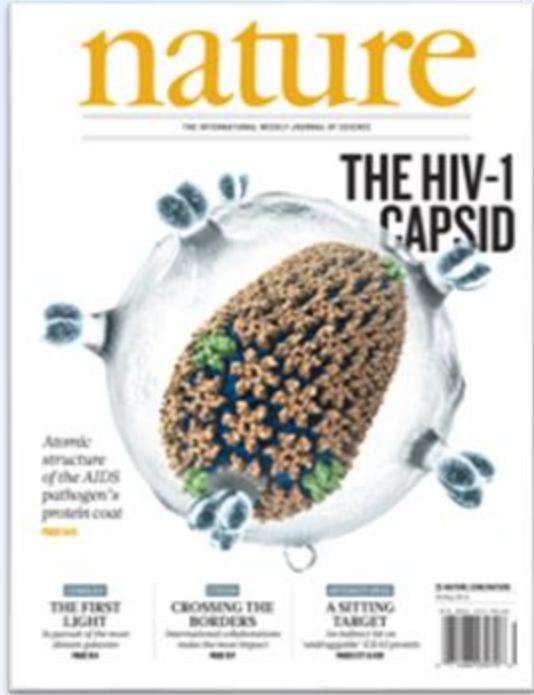
100-300 Gbps WAN



Spectra Logic: 300 usable PB



Soneion: 26 usable PB

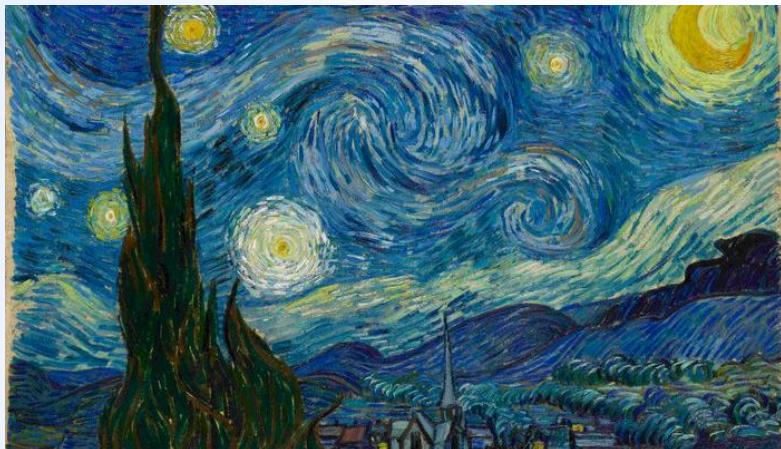


136M atoms! AIDS Virus, Chromatophore
Schulten, UIUC

UNDERSTANDING LIFE &
OTHER MATERIALS!

Starry Night

Turbulence in Fluids



Vincent van Gogh, 1889

“When I die and go to heaven there are two matters on which I hope for enlightenment. One is quantum electrodynamics, and the other is the turbulent motion of fluids. And about the former I am rather optimistic.” Horace Lamb, 1932...

- Turbulent flow via direct numerical simulation (DNS) just possible on modern Petascale systems
 - 100PF+ would allow complete engine simulation with turbulence
 - 40 Exaflop would allow complete airplane simulation!

CONVERGENCE OF ARTS, COMPUTING AND DATA SCIENCES



Lifetime Achievement Award , IMERSA 2017 Summit, for

“Significant contributions to the world of full dome film. This honor is truly the highest accolade we can give to someone who has had such a profound impact on our profession.”

Donna Cox

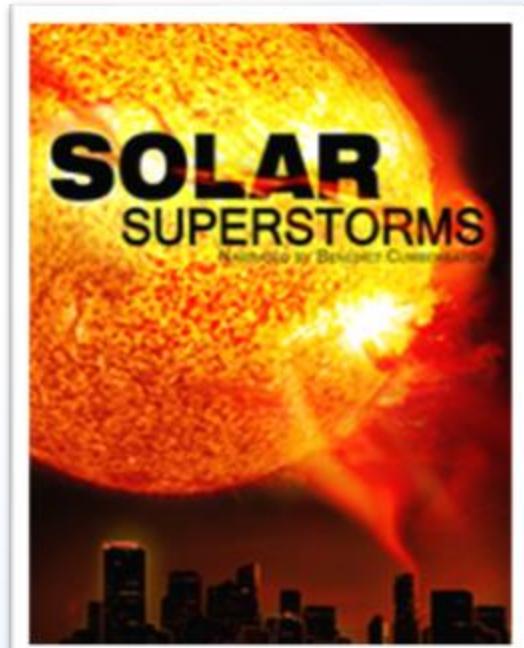
Michael Aiken Chair

Professor, School of Art and Design

Director, Illinois Emerging Digital Research and

Education in Arts Media Institute (eDream)

Head, Research and Education Directorate, NCSA



IMAX Hubble 3D narrated by Leonardo DiCaprio



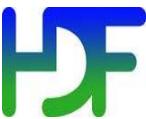
Manufacturing



Rolls-Royce



Technology



CRAY



allinea



Charmworks



DataDirect
NETWORKS™



Energy

ExxonMobil

Upstream Research



MAYO
CLINIC



AGRIBLE

syngenta

Life Sciences



THE DARK ENERGY SURVEY



Manufacturing

Technology

Institutes

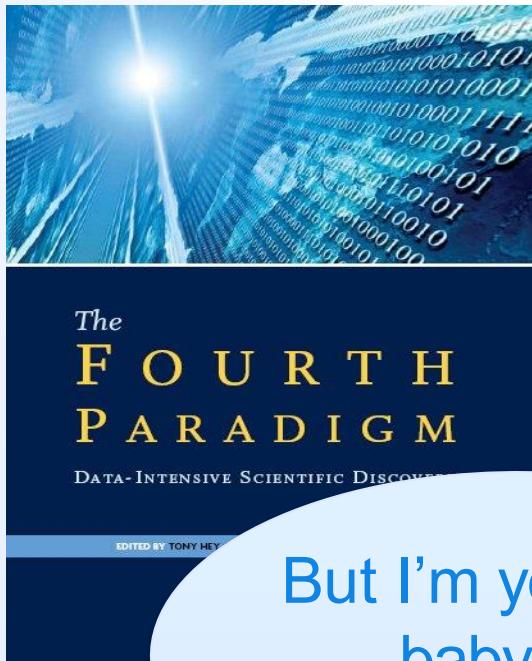
Half of the Blue Waters
Projects Required
Connection to Real-
World Data...

New Machines nearly
500x as powerful aim
to integrate Data, AI,
Experiment!

Real World Science and Engineering Require
Integration of Compute and Data

FUTURE DIRECTIONS

The Growth of Data



But I'm your new
baby big
brother...



Data Tsunami



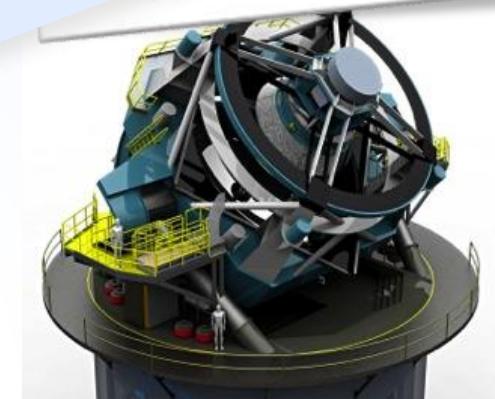
With millions of
processors...

"I'm still
here..."

Data-enabled Transformation of Science



Kathy Gray: How can I publish, discover, verify data in this new world?



Astronomy 1500- 2000:

- Single scientist looks through telescope
- Record KB of data in notebook
- Require reproducibility

Sloan Digital Sky Survey 2000+

- Record data for decade (40TB)
- Serve to entire world
- Thousands of scientists work “together”

DES (2013-19)

- 200GB/night
- PB in decade

LSST (Rubin: 1 year)

- Record data for decade
- SDSS/night!
- 200 PB/decade

Integrating it all: Multi-Messenger Astronomy...

Astronomy c. 2025!

- New era: seeing events as they occur
- Here now
 - ALMA, EVLA
 - Ice Cube neutrinos
- On horizon
 - 20-30m optical?
 - LSST
 - aLIGO, Indigo
- SKA = exabytes
- Simulations integrate physics



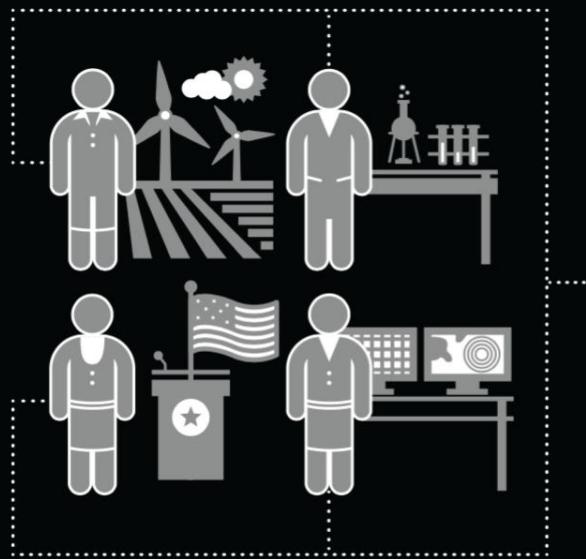
Communities communicate by sharing data, software... knowledge...

...Or Sustainable Food Production...



Complex Grand Challenges: An Example

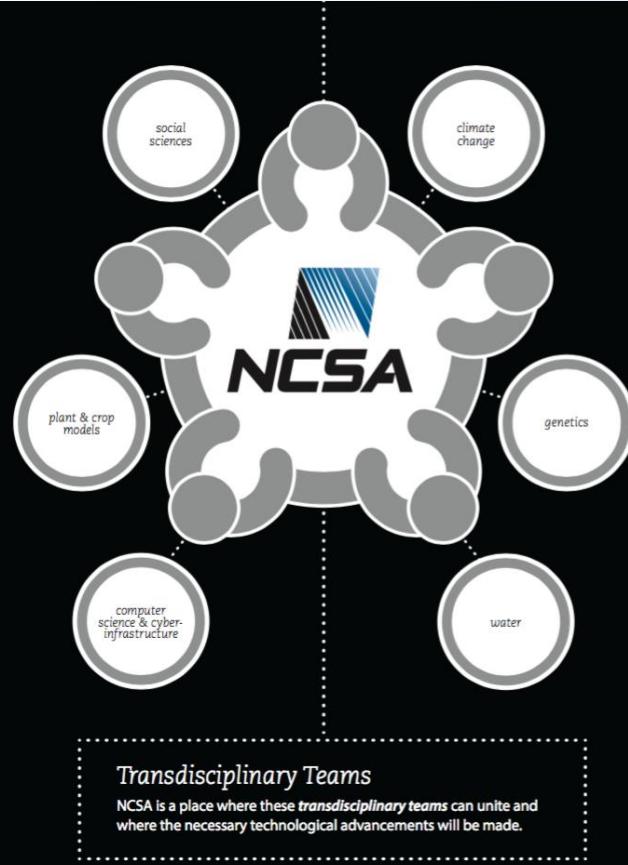
The U.S. Midwest is the largest food producer in the world. Current typical agricultural practices have a detrimental impact on the Mississippi Basin and Gulf of Mexico, as fertilizer runoff causes nitrogen hypoxia—essentially, leaving a large “dead zone” in the Gulf. Extensive data on land usage, fertilizer runoff, water flows, crop genetics, and many other factors affecting the growth of crops and its environmental impacts are being collected, and need to be interpreted and acted upon.



Sustainable Food Growth

How can U.S. farmers best grow plants to feed the world while reducing negative environmental impacts?

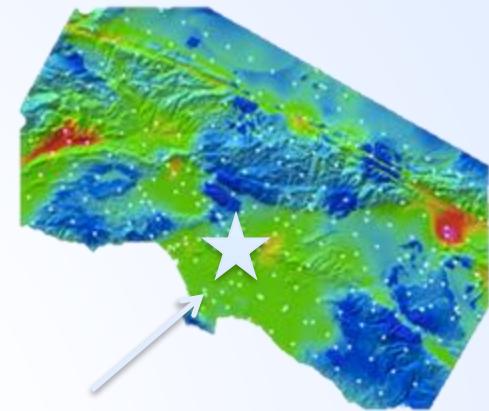
This is one motivating grand challenge problem for NCSA’s thematic areas, crossing *Earth and Environment, Bioinformatics and Health Sciences, Culture and Society, and Computing and Data Sciences*. This challenge and others like it require expertise and engagement from multiple disciplines and also demand fundamental advances in the underlying methodologies of how we derive meaning from massive quantities of data and how we integrate modeling of phenomena at multiple scales.



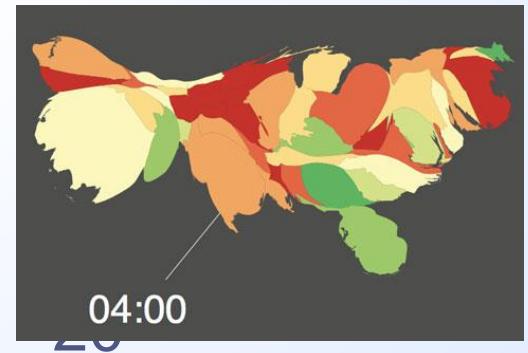
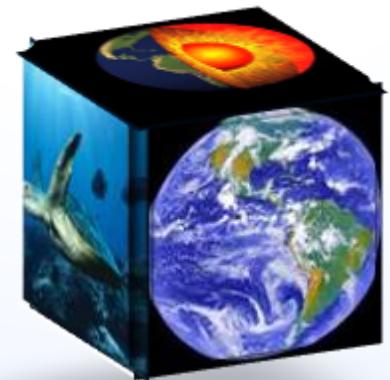
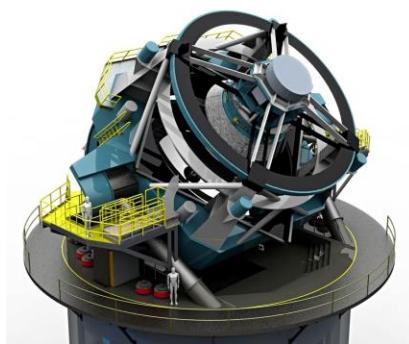
Transdisciplinary Teams

NCSA is a place where these *transdisciplinary teams* can unite and where the necessary technological advancements will be made.

Scenarios like this in all fields...



Los Angeles



Big Data vs The Long Tail of Science

- Many “Big Data” projects are “special”
 - Tend to be highly organized, have singular sources of data, professionally curated, a lot of attention paid to them
- What about the “Long Tail” (the other 99%)?
 - Thousands of biologists sequencing communities of organisms
 - Thousands of chemist and materials scientists developing a “materials genome”
 - Millions of people “Tweeting”...
 - Characteristics:
 - Heterogeneous, perhaps hand generated
 - Not curated, reused, served, etc...





Kathy Gray: How can
I publish, discover,
verify data in this
new world?

But we need them...

**STILL STUMBLING TO CREATE
UBIQUITOUS DATA SERVICES**

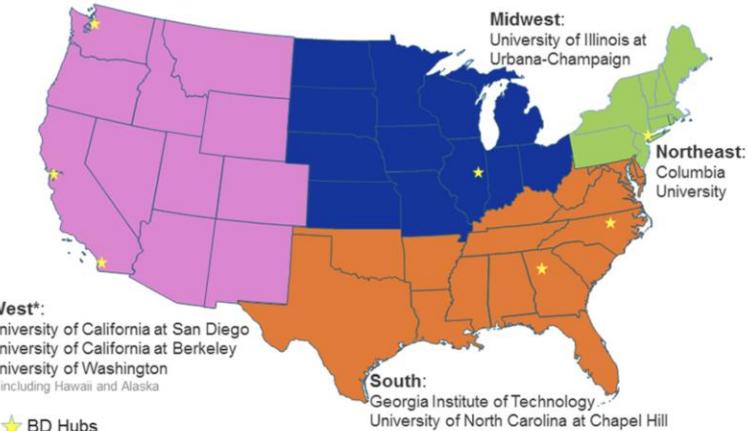
Basic Vision for Open Services

- Make it possible
 - Create a service
 - Deposit it in a service
 - Provide services to repurpose it...
 - Link it to traditional publications
 - OA aspects very important to this
- With these capabilities in place
 - Many important things will happen...
 - Please:) publications...

"We need to take steps to make scientific research data more liquid. The more we move towards open as the default for scientific research data, the more we will get out of the research enterprise. It is time to take deliberate steps to make that a reality." Mike Stebbins, White House OSTP

Open, Shareable Data: Critical for the future

- *Interdisciplinarity and complex problem solving*
 - Needed: ability to find, integrate results across communities
- *Reproducibility of a scientific result*: heart of science
 - Needed: access to complete state of a result, including data, software, methods, (and the publication itself)
- *Accelerating discovery*: faster, deeper dissemination of results to other researchers; *Repurposing data* by others: extending in new ways
 - Needed: services to find, retrieve, analyze, describe data/results
- *Economic development*
 - Needed: availability of all the above to companies (MGI!)
- *Public dissemination* of publicly funded research results
 - Needed: open, accessible results, searchable by public


[Home](#)
[About this Workshop](#)
[Agenda](#)
[Hotel & Registration](#)

Big Data Regional Innovation Hubs and Spokes Workshop #BDHubs

Data services

SOME ATTEMPTS TO GET GOING...



The National DATA SERVICE

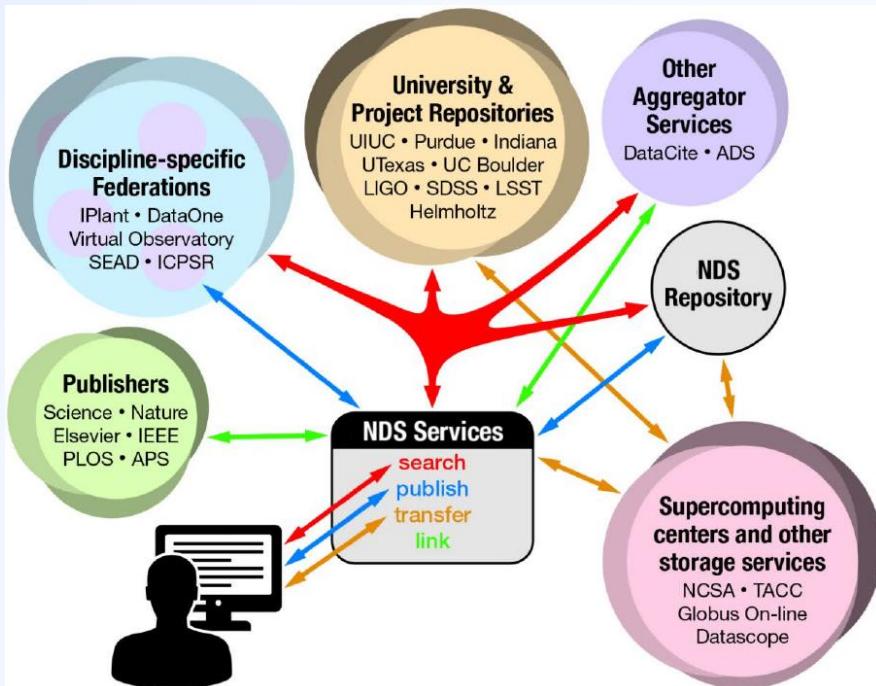
National Data Service Workshop
October 19-21, 2015



1st National Data Service Consortium Workshop

June 12-13, 2014 Boulder, Colorado

This meeting is the first gathering of stakeholders interested in establishing a National Data Service and the Consortium that can guide its development.



The National DATA SERVICE

National Data Service Workshop
October 19-21, 2015

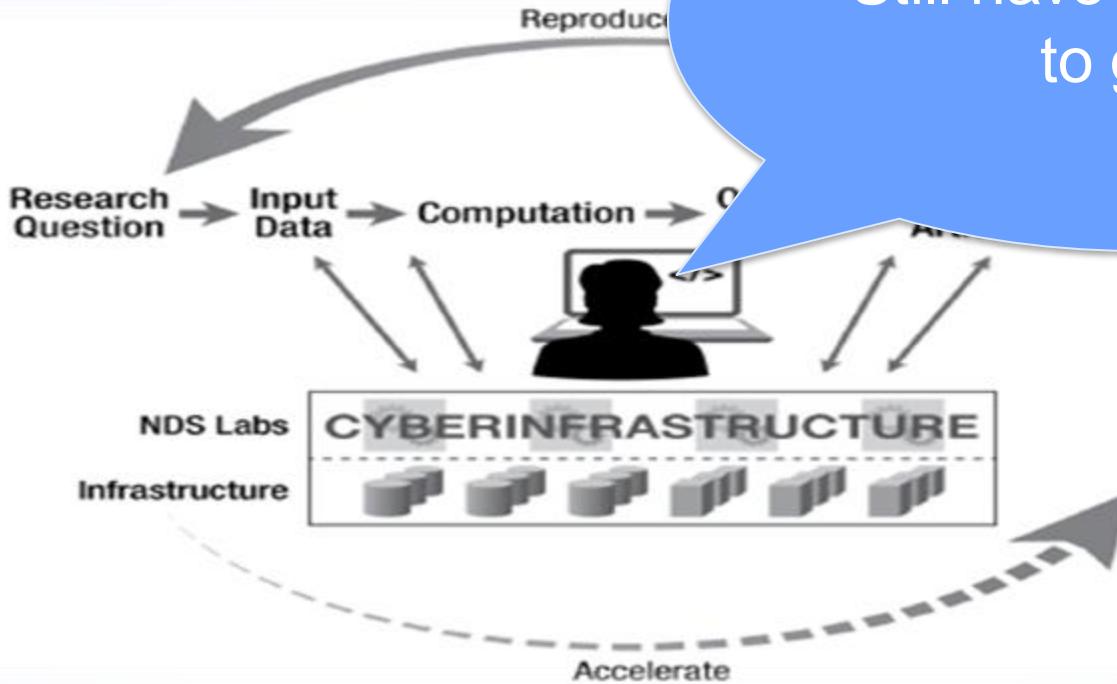
The National Data Service Consortium fourth plenary meeting will be in San Diego, October 19-21 and limited space is still available!

Researchers, educators, students communicate by sharing data
Services needed to make it work!

- To create infrastructure and services that truly enables promise of data-science
 - National scale, but federated down to individual campuses and repositories
- NDS should help researchers...
 - Find data...
 - Use Data...
 - Share and publish data...
- Extend/integrate efforts of individual projects

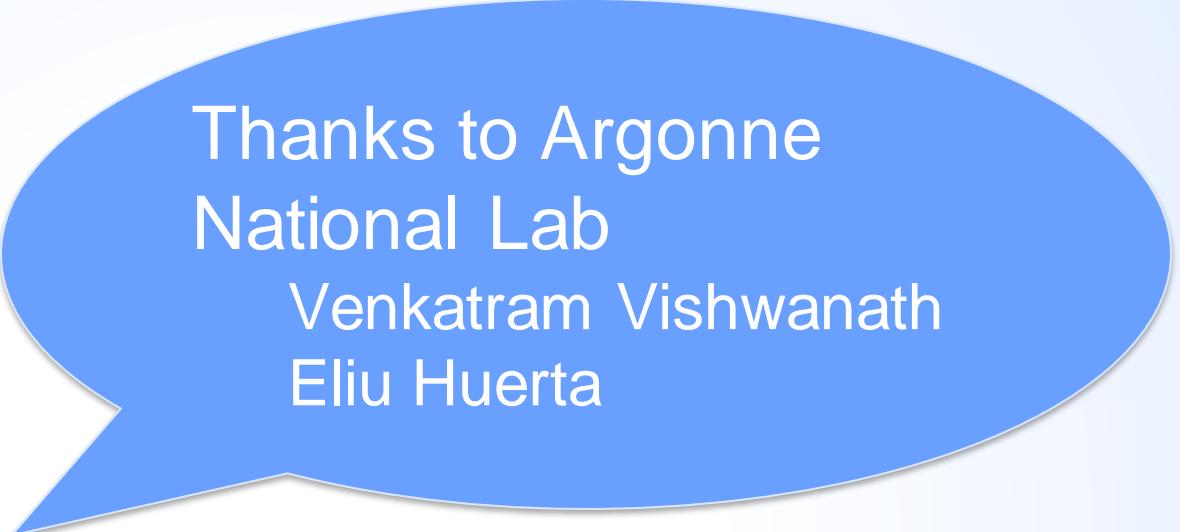
The Whole Tale Project (WTP)

Still have a long way to go...



- iPython notebooks, DOIs, linked publications, data, tools
- Support Open Science principles for reproducible, extensible research and building science communities
- Working with National Data Service, publishers
 - Science, BioOne, APS, ...





Thanks to Argonne
National Lab

Venkatram Vishwanath
Eliu Huerta

AI and Exascale

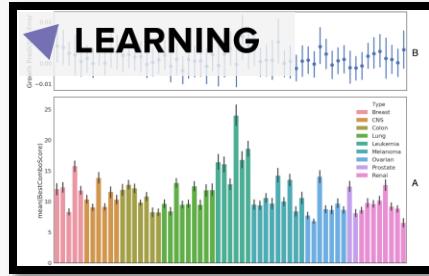
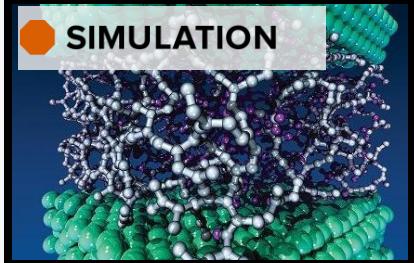
**MAJOR PROGRESS
RECENTLY**

Argonne Leadership Computing Facility



The Argonne Leadership Computing Facility provides world-class computing resources to the scientific community.

- Users pursue scientific challenges
- In-house experts to help maximize results
- Resources fully dedicated to open

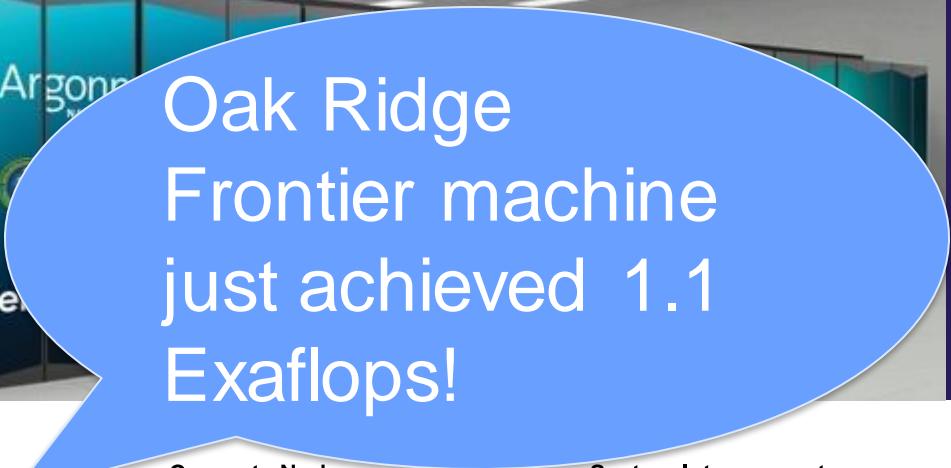


ALCF offers different pipelines based on your computational readiness. Apply to the allocation program that fits your needs.



Architecture supports three types of computing

- § Large-scale Simulation (PDEs, traditional HPC)
- § Data Intensive Applications (scalable science pipelines)
- § Deep Learning and Emerging Science AI (training and inferencing)



Oak Ridge Frontier machine just achieved 1.1 Exaflops!

PEAK PERFORMANCE

≥ 2 Exaflops DP

Intel GPU

Ponte Vecchio

Intel Xeon PROCESSOR

Sapphire Rapids wt
HBM

PLATFORM

HPE Cray-Ex

Compute Node

2 SPR+HBM processor;
6 PVC; Unified
Memory Architecture; 8 fabric
endpoints;

GPU Architecture

Xe arch-based "Ponte Vecchio"
GPU
Tile-based chiplets
HBM stack
Foveros 3D integration

System Interconnect

HPE Slingshot 11; Dragonfly
topology with adaptive routing

Network Switch

25.6 Tb/s per switch, from 64–200
Gb/s ports (25 GB/s per direction)

Node Performance

>130 TF

System Size

>9,000 nodes

Aurora

Leadership Computing Facility
Exascale Supercomputer

Aggregate System Memory

>10 PB aggregate System Memory

High-Performance Storage

220 PB @ EC16+2, ≥25 TB/s DAOS

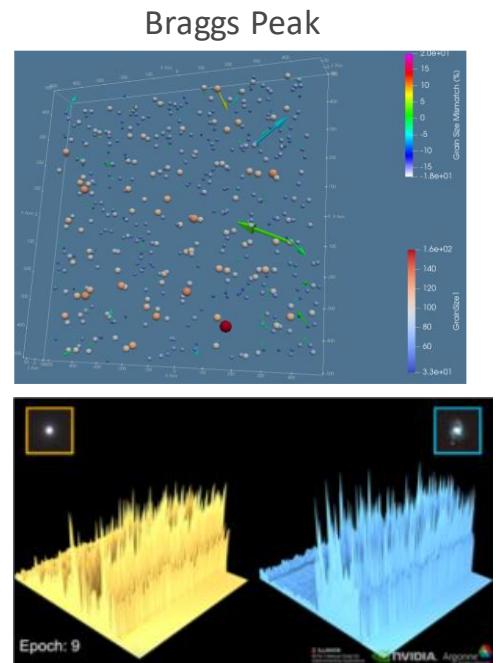
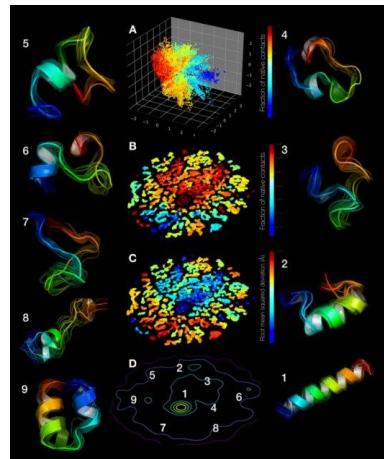
Programming Models

oneAPI, MPI, OpenMP, C/C++,
Fortran, SYCL/DPC++
Python-based environments
Machine learning and Deep learning
frameworks

SURGE OF SCIENTIFIC MACHINE LEARNING

- Simulations/ surrogate models
 - Replace, in part, or guide simulations with AI-driven surrogate models
- Data-driven models
 - Use data to build models without simulations
- Co-design of experiments
 - AI-driven experiments

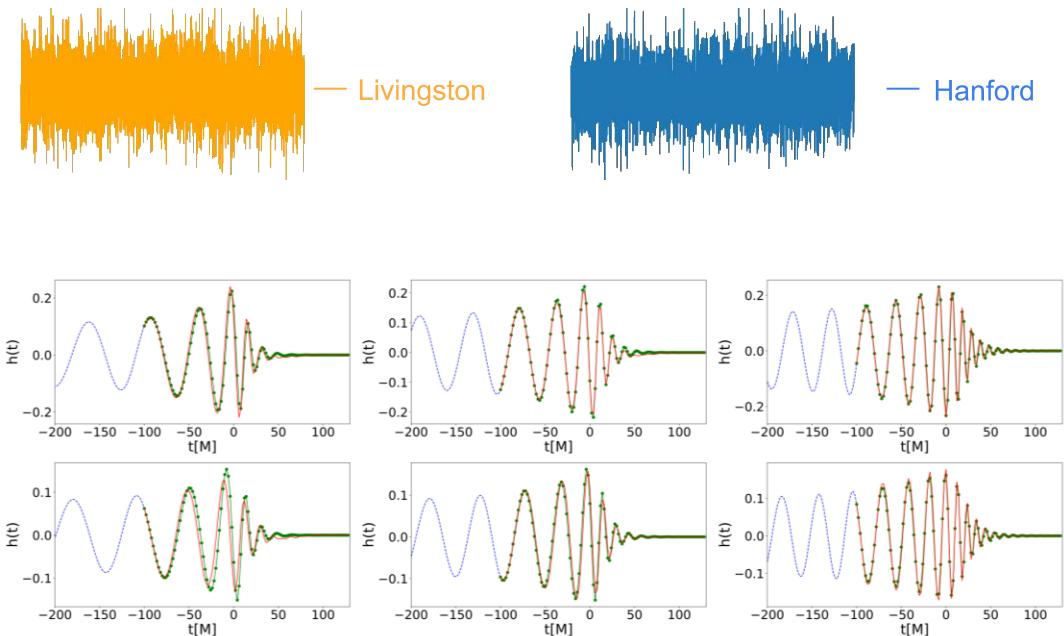
**Design infrastructure to facilitate and accelerate
AI for Science applications**



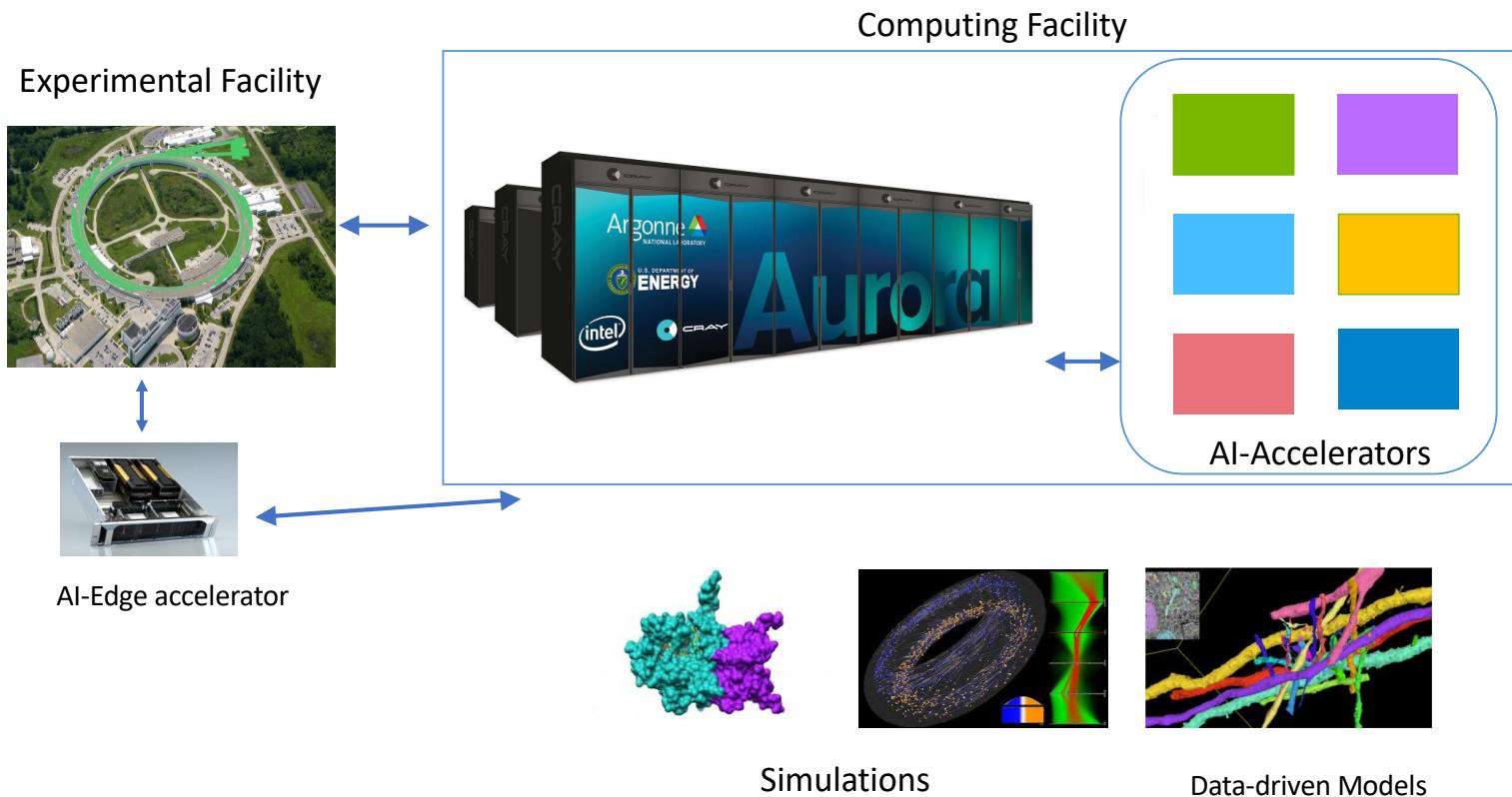
SAMPLE CASE: GRAVITATIONAL WAVE ASTROPHYSICS

How

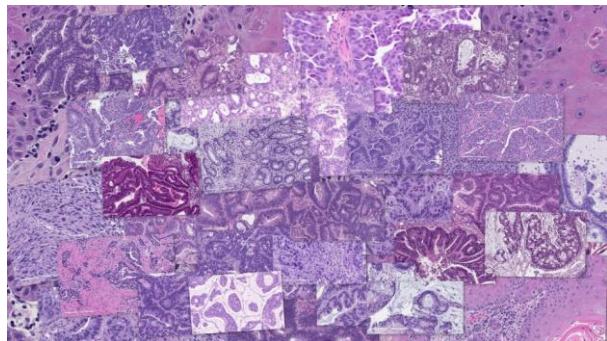
Grand challenge: identify weak signals embedded in large backgrounds, experimental noise is non-Gaussian and non-stationary



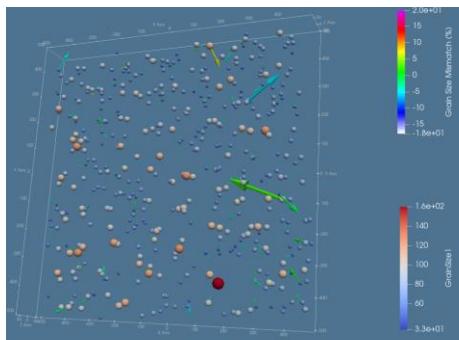
INTEGRATING AI SYSTEMS IN FACILITIES



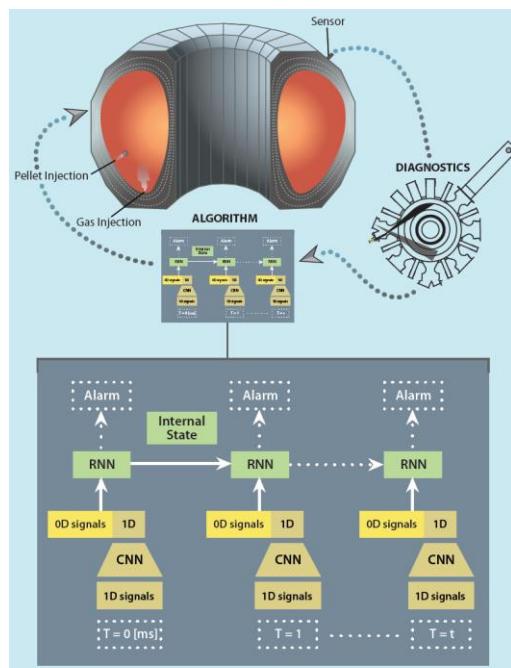
AI FOR SCIENCE APPLICATIONS ON AI TESTBED



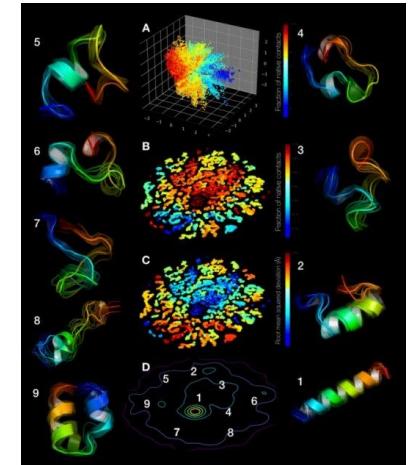
Cancer drug response prediction



Imaging Sciences-Braggs Peak



Tokomak Fusion Reactor operations



Protein-folding (Image: NCI)

and more..

Bringing it all to Mountain West Region

- Can share data, computing, expertise
 - Pools of resources
 - Regional capacity
- Wyoming Developments
 - University of Wyoming School of Computing
 - Computing applications across the spectrum
 - Advanced Research Computing Center (ARCC)
 - New Director Liuda Mainzer (Formerly NCSA)
 - NCAR-Wyoming Supercomputing Center
- New opportunities
 - Chips and Science Act!
 - NSF Technology, Innovation, and Partnerships Directorate
- Let's put some ideas together!

Summary

- Major paradigm and culture shifts: research is changing dramatically
 - Complex problems require collaborations at new scales
 - Computing and data capabilities growing at unprecedented rate
- Major infrastructure projects are highly computing & data intensive
 - Data services are needed to support
 - A comprehensive, integrated approach would better serve science, reduce costs
- Lots of opportunities in our region now!

