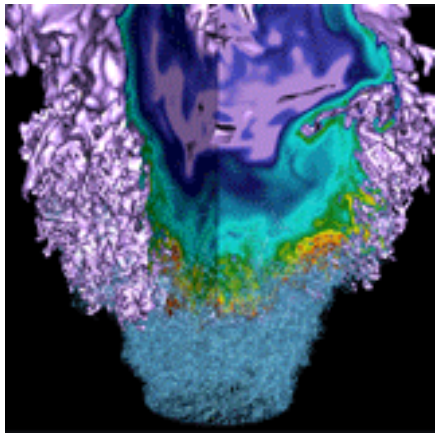
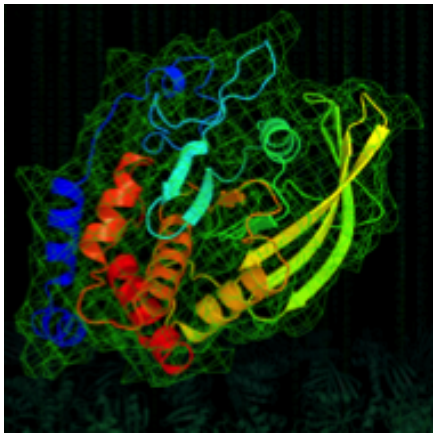
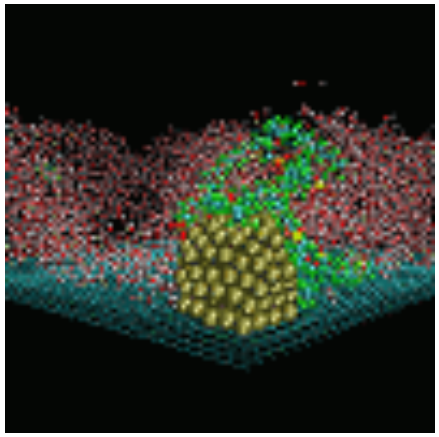
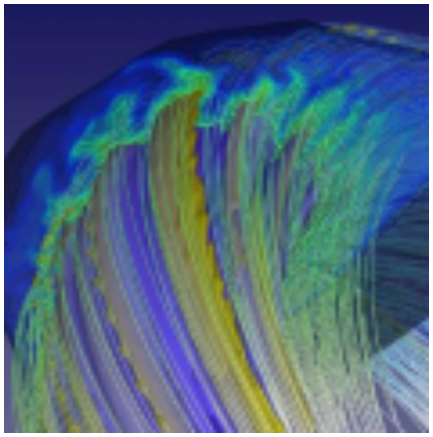
# Stories at NERSC

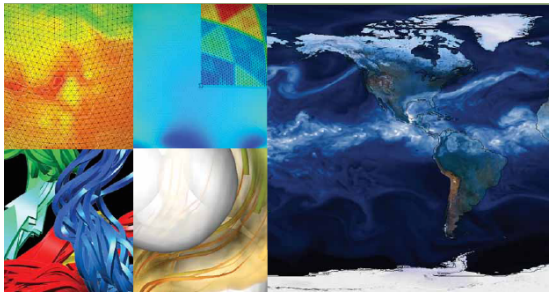






Joaquin Correa
Data Analytics Services

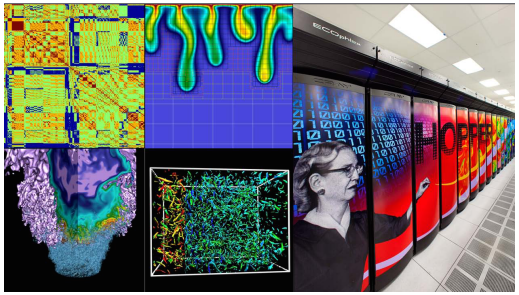
# NERSC is the Mission HPC Facility for DOE Office of Science Research

U.S. DEPARTMENT OF ENERGY | Office of Science

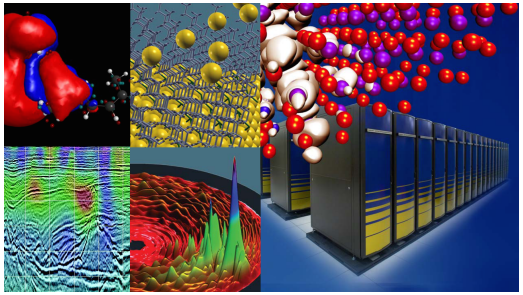
Largest funder of physical science research in U.S.



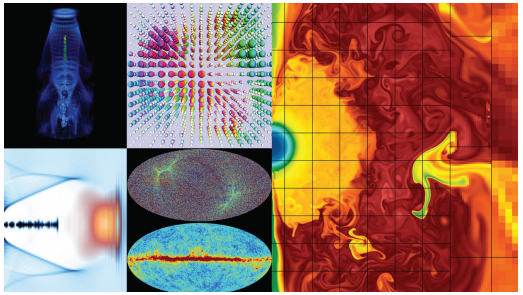
Bio Energy, Environment



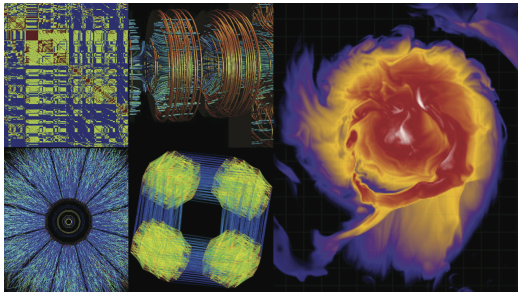

Computing



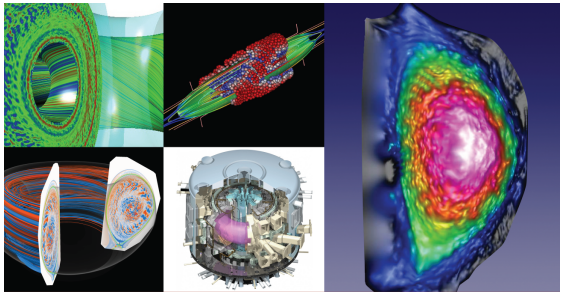
Materials, Chemistry, Geophysics



Particle Physics, Astrophysics



Nuclear Physics



Fusion Energy, Plasma Physics

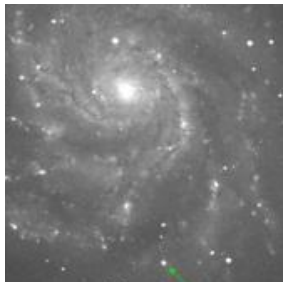
# NERSC deploys first of a kind HPC systems for Department of Energy mission science

- We collaborate with computer companies years before a system's delivery to deploy advanced systems with new capabilities at large scale
- We provide a highly customized software and programming environment for science applications
- We are tightly coupled with the workflows of DOE's experimental and observational facilities – ingesting tens of terabytes of data each day
- Our staff provide advanced application and system performance expertise to users


# NERSC Measures itself by the Science it Enables

## Astrophysics

NERSC played a key role in the discovery that led to the 2011 Nobel Prize in Physics.
**(S. Perlmutter, UC Berkeley/LBNL)**

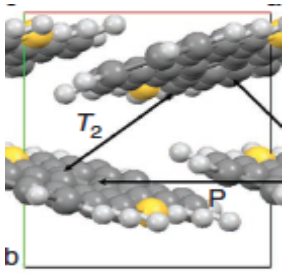


## Astrophysics

The earliest-ever detection of a supernova was made possible by NERSC and Esnet.
**(P. Nugent, LBNL)**

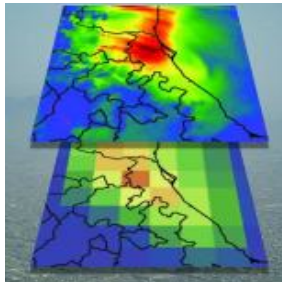
## Materials

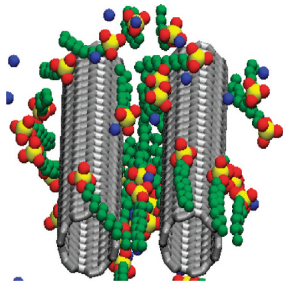
A vastly improved organic semiconductor discovery is a key proof of principle for rational design of new materials.
**(A. Aspuru-Guzik, Harvard)**




## Climate

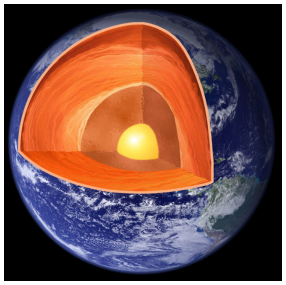
Atmospheric scientists have shown how small-scale effects of aerosols contribute to errors in climate models.
**(W. Gustafson, PNNL)**





## Chemistry

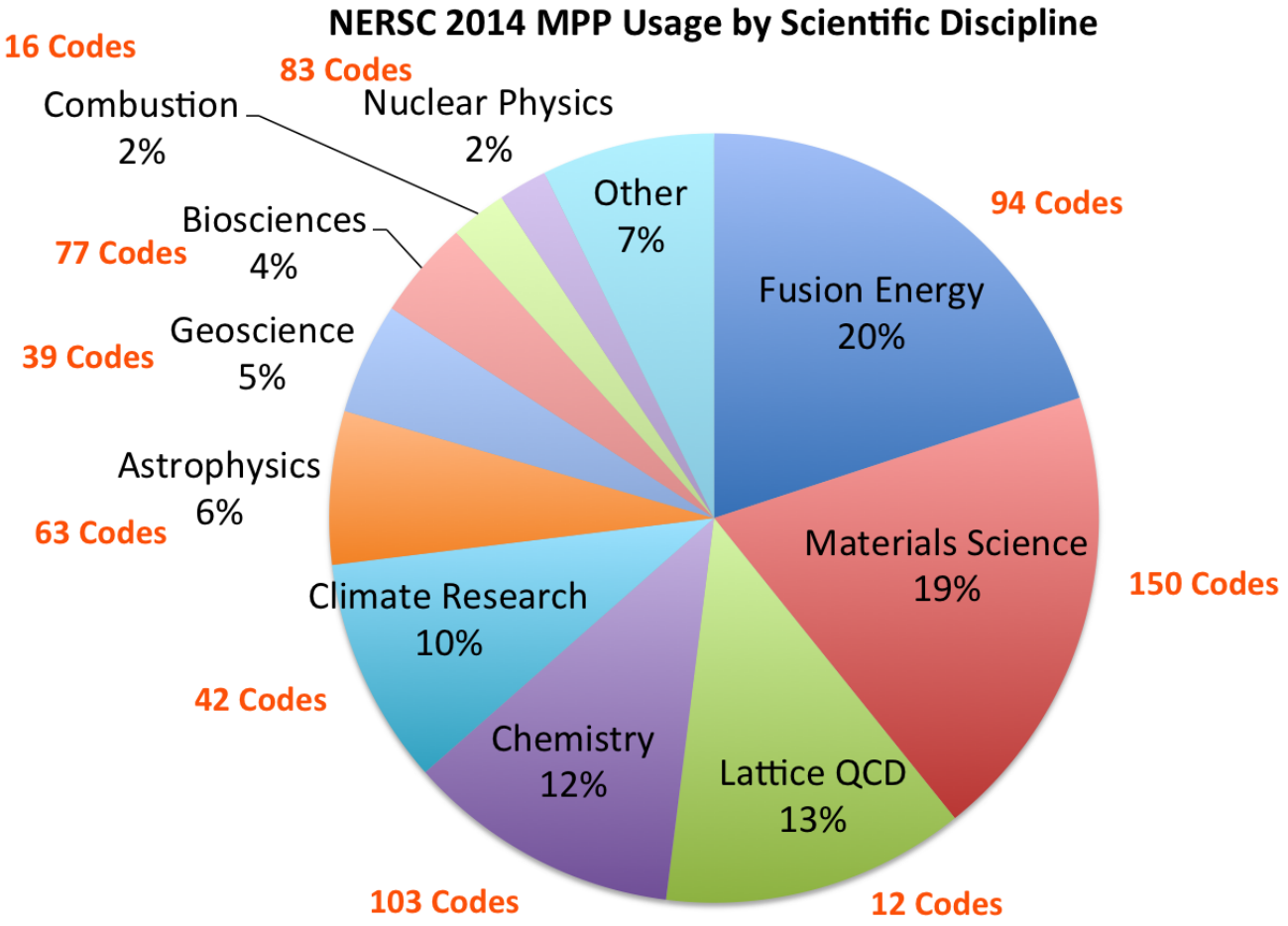
Molecular dynamics simulations show how certain surfactants can be used to separate out bundles of carbon nanotubes with important properties.
**(A. Striolo, U. Oaklahoma)**



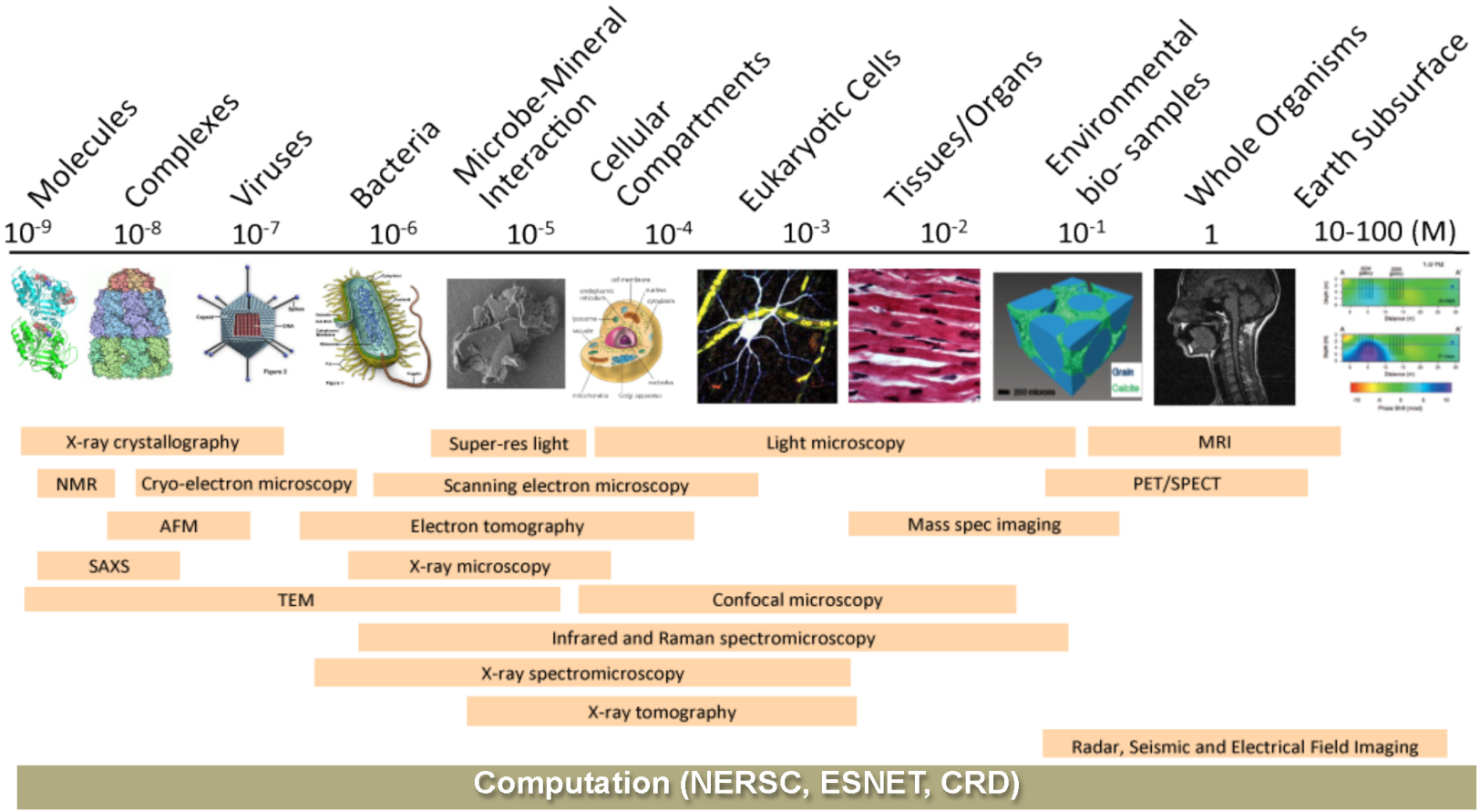
## Nuclear Physics

The KamLAND neutrino experiment showed that radioactivity cannot be Earth's only heat source; it accounts for only ½ of it.
**(S. Freedman, LBNL)**

# Computing Usage at NERSC 2014

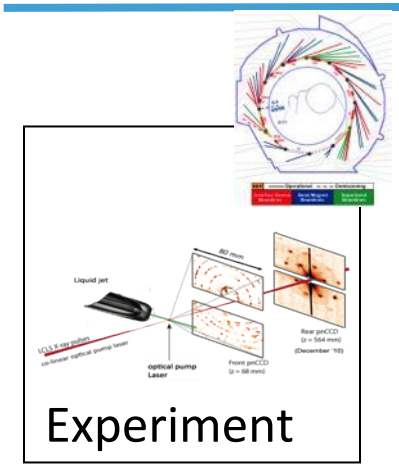


# Computation and Data are Crosscutting




| | |
|---|---|
| Data driven high throughput phenomenology<br>knowledge networks, new models | Image analysis<br>Machine learning |

# Beamline Scientific Workflow

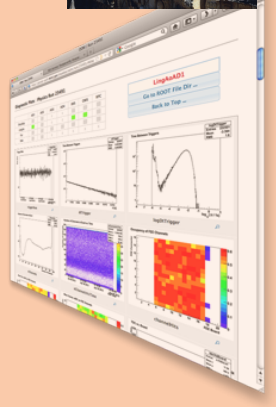



Experiment

Data Pipeline

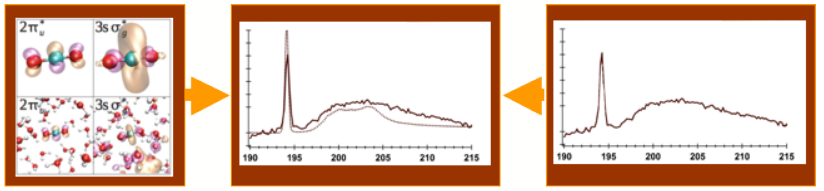
HPC Storage and Compute


Beamline User



Science Gateway

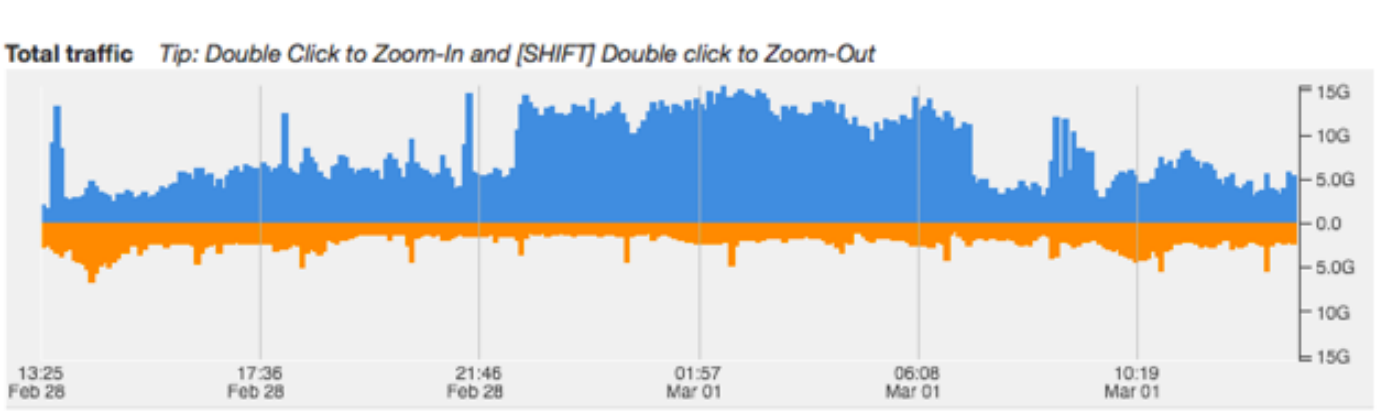
simulate → compare ← measure




Simulation and Analysis Framework

# LCLS -> NERSC

All NERSC Traffic




Photosystem II X-Ray Study

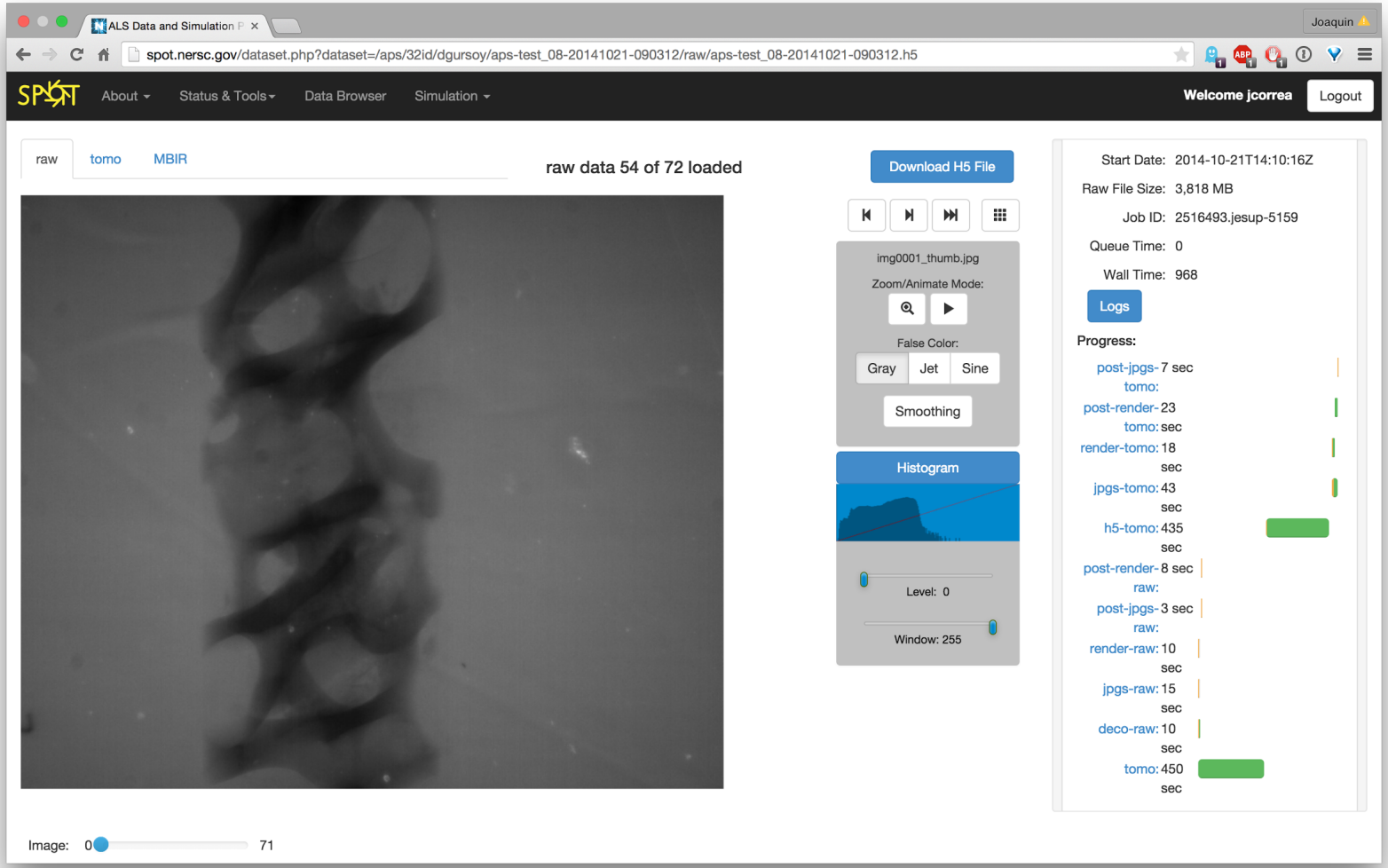



ESnet Energy Sciences Network

U.S. DEPARTMENT OF ENERGY | Office of Science

BERKELEY LAB

# Reengineering of SPOT using OMERO



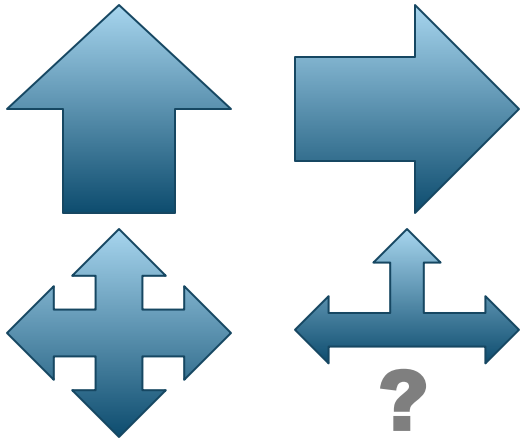
# Thank you!

# Big Data and Big Computing Need Each other

- **Big data requires simulation**
  - Missing data; models of what the data means
- **Big data requires big computing**
  - Large data requires large computing (and memory and storage)
- **Big computing generates big data**
  - Petaflops in many science areas
- **Big Data benefits science**
  - More science for investments

# What is Big Data? : Attention to Data

- **When data focus is <u>unavoidable</u> or science is "<u>in the data</u>"**
- **4Vs : facets of "Big" data**
  - Volume (non-trivial sizes)
  - Velocity (fast, bandwidth)
  - Variety (complex, connected)
  - Variability (gaps, errors)
- **Sources/causes of data focus?**
  - Increasing detector resolutions
  - Increasing repetition rates
  - High throughput screening methods



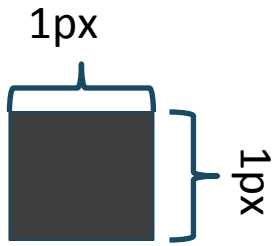
# Images in HPC and Data-centric Science

## Not new

- **Telescopes → surveys**
  - FITS, lots of metadata
  - Analysis pipelines from instrument to HPC
  - Register, overlay, analyze, share
- **Gridded simulation data is often block/raster oriented**
- **Data at scale**
- **Reliable computing at scale**
- **Goal is knowledge discovery (Knet, actionable models)**

## Newly interesting

- **Drivers**
  - Detectors
  - Automation
  - Collaboration, Crowds
  - Data Policies
- **Meet the enemy**
  - The pixel: not a great data model




- **Key challenge**
  - Can HPC speed the way from pixels to knowledge?