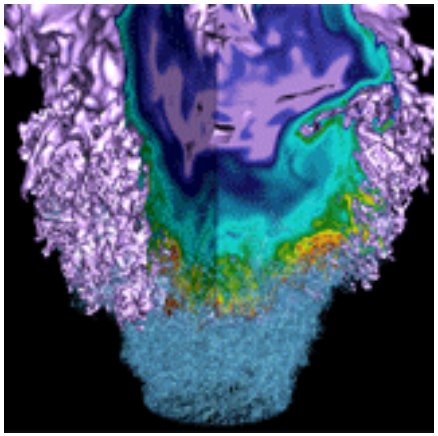
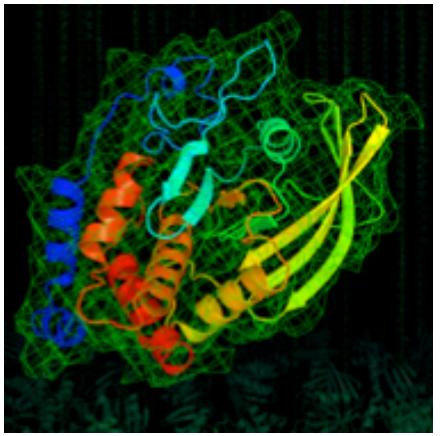
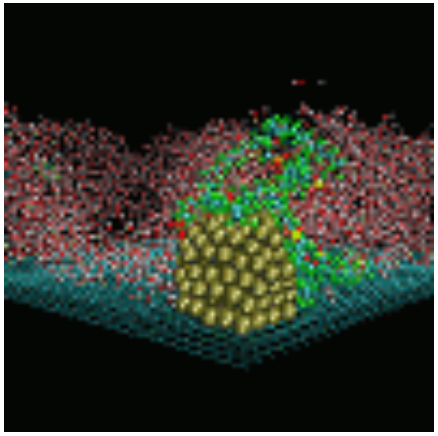
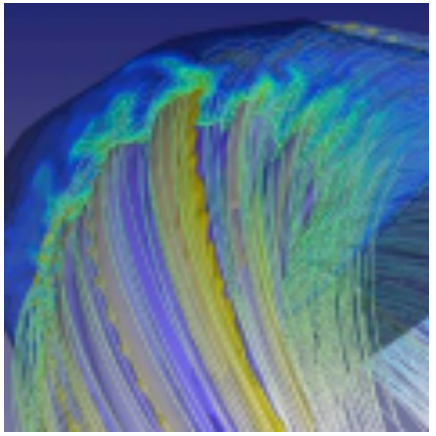
Integrated tool for next generation bio-imaging

Cryo-EM Meeting
Stanford University/SLAC National Laboratory

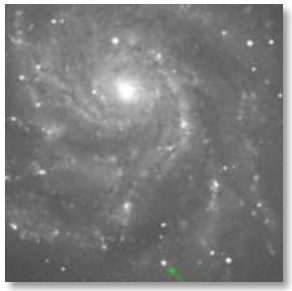
Joaquin Correa
Data & Analytics Group

April 2014

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)

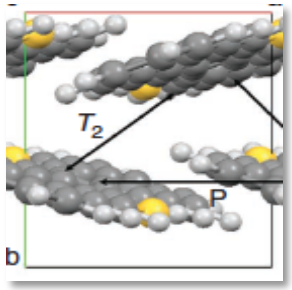


Astrophysics

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

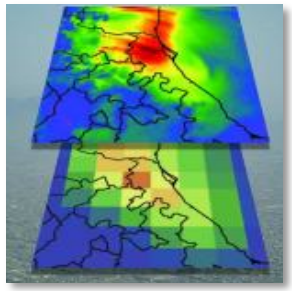
Materials

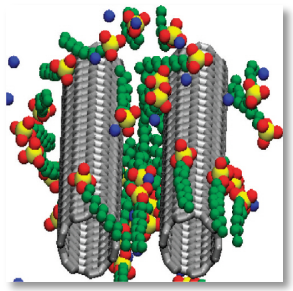
A vastly improved organic semiconductor discovery is a key proof of principle for rational design of new materials.
(A. Aspuru-Guzik, Harvard)



Climate

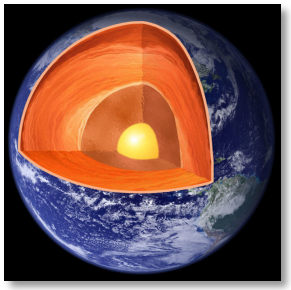
Atmospheric scientists have shown how small-scale effects of aerosols contribute to errors in climate models.
(W. Gustafson, PNNL)





Chemistry

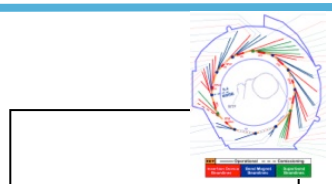
Molecular dynamics simulations show how certain surfactants can be used to separate out bundles of carbon nanotubes with important properties.
(A. Striolo, U. Oaklahoma)



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)

ALS Scientific Workflow



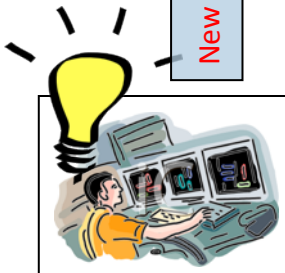
Experiment

Data Pipeline

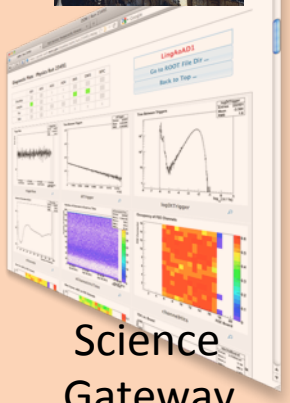
HPC Storage and Compute

Prompt Analysis Pipeline

New experiment

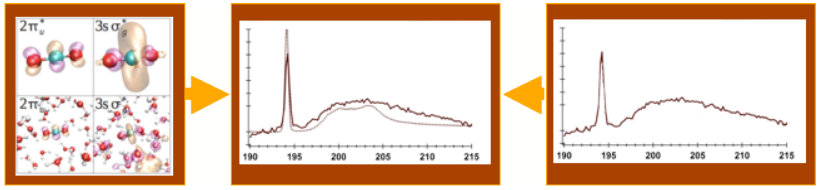


Beamline User



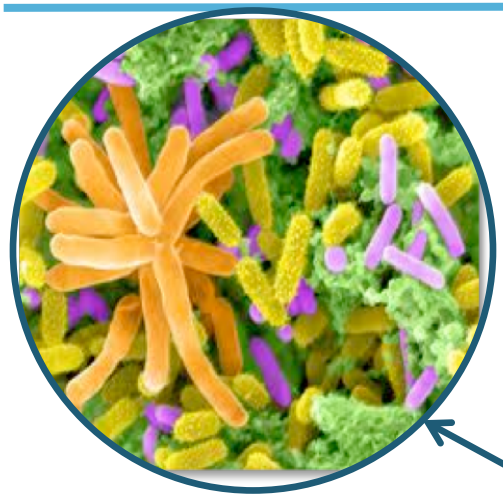
Science Gateway

simulate → compare ← measure



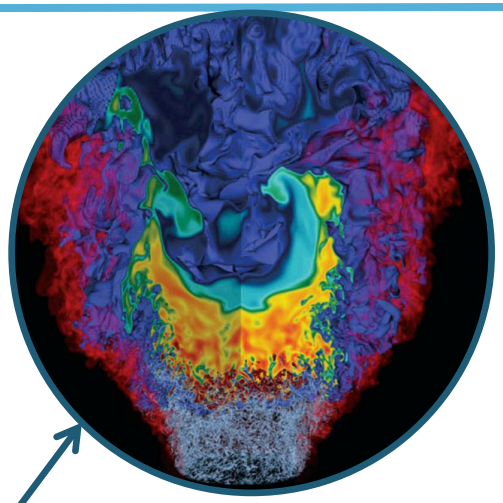
Simulation and Analysis Framework

Computing is Essential for Science Programs in All Areas of the Lab

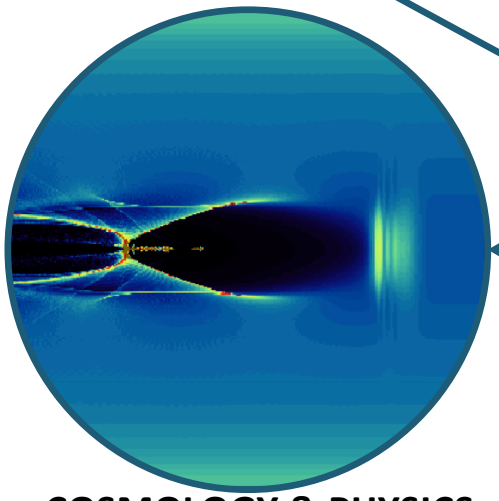


BIOSCIENCES

ENERGY & ENVIRONMENTAL



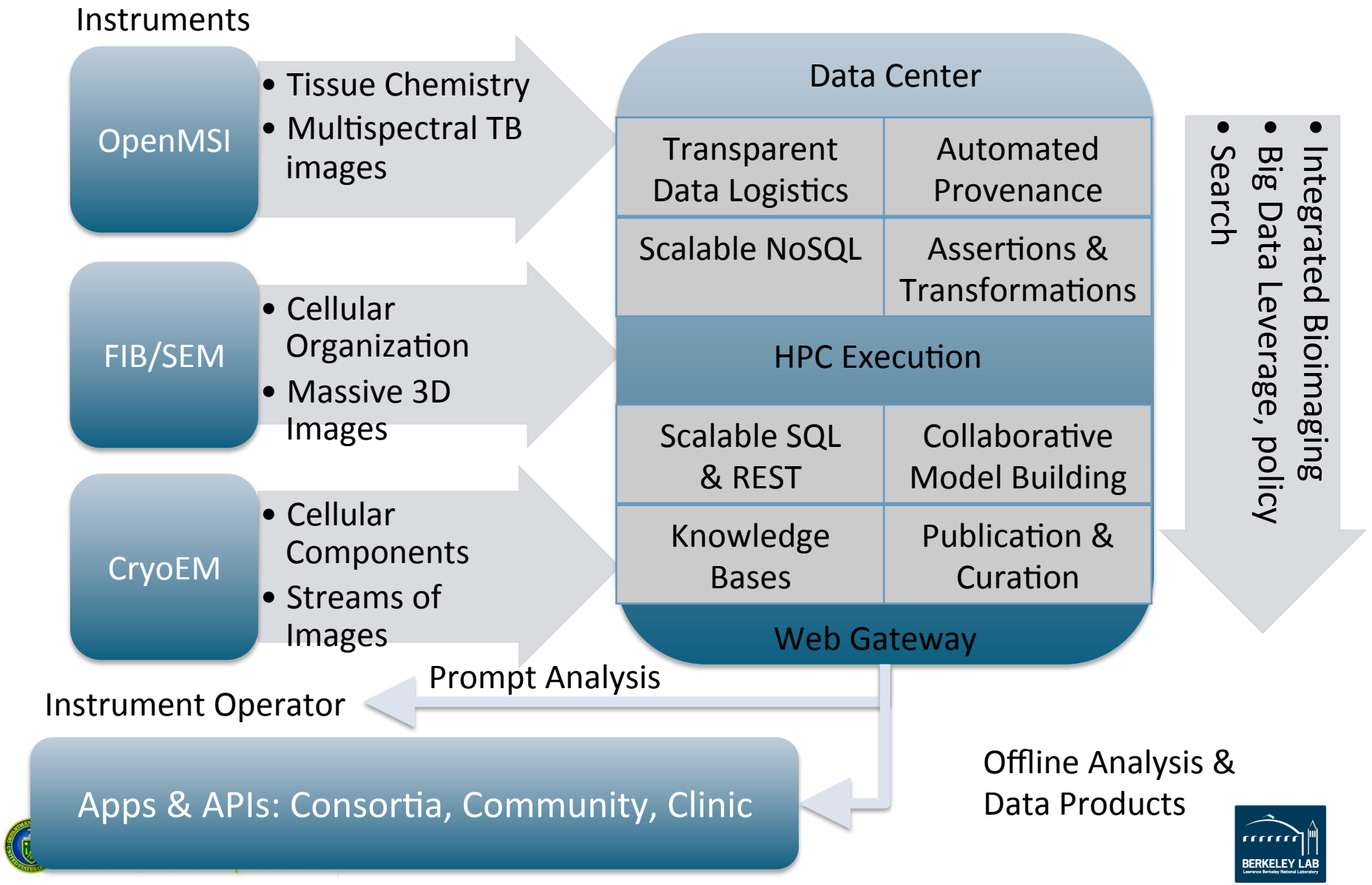
PHOTON SCIENCES



COSMOLOGY & PHYSICS

COMPUTING SCIENCES

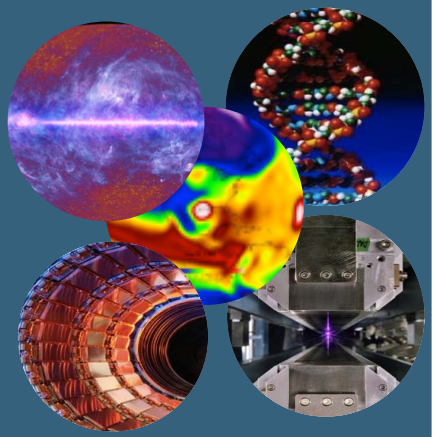
Big Data Bioimaging Workflow



Simulation and analysis use high end computing via a science-capable Network

Big Data

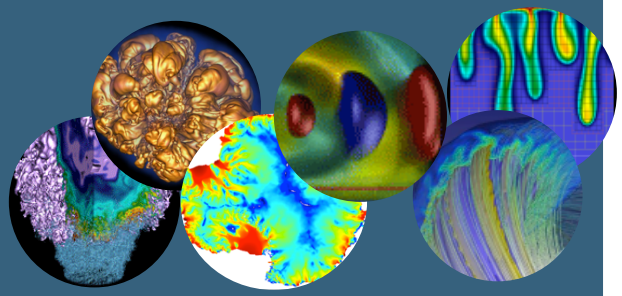
From Experiments and Simulation



NERSC ingests, stores and analyzes data from Telescopes, Sequencers, Light sources, Particle Accelerators (LHC), climate, and environment

Large Scale

Capability Simulations



Petascale systems run simulations in Physics, Chemistry, Biology, Materials, Environment and Energy at NERSC

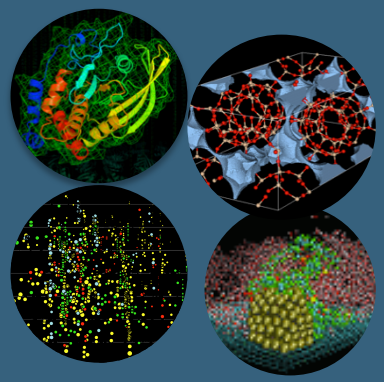
NERSC

Petascale Computing, Petabyte Storage, and Expert Scientific Consulting

ESnet
Energy Sciences Network

High Volume

Job Throughput

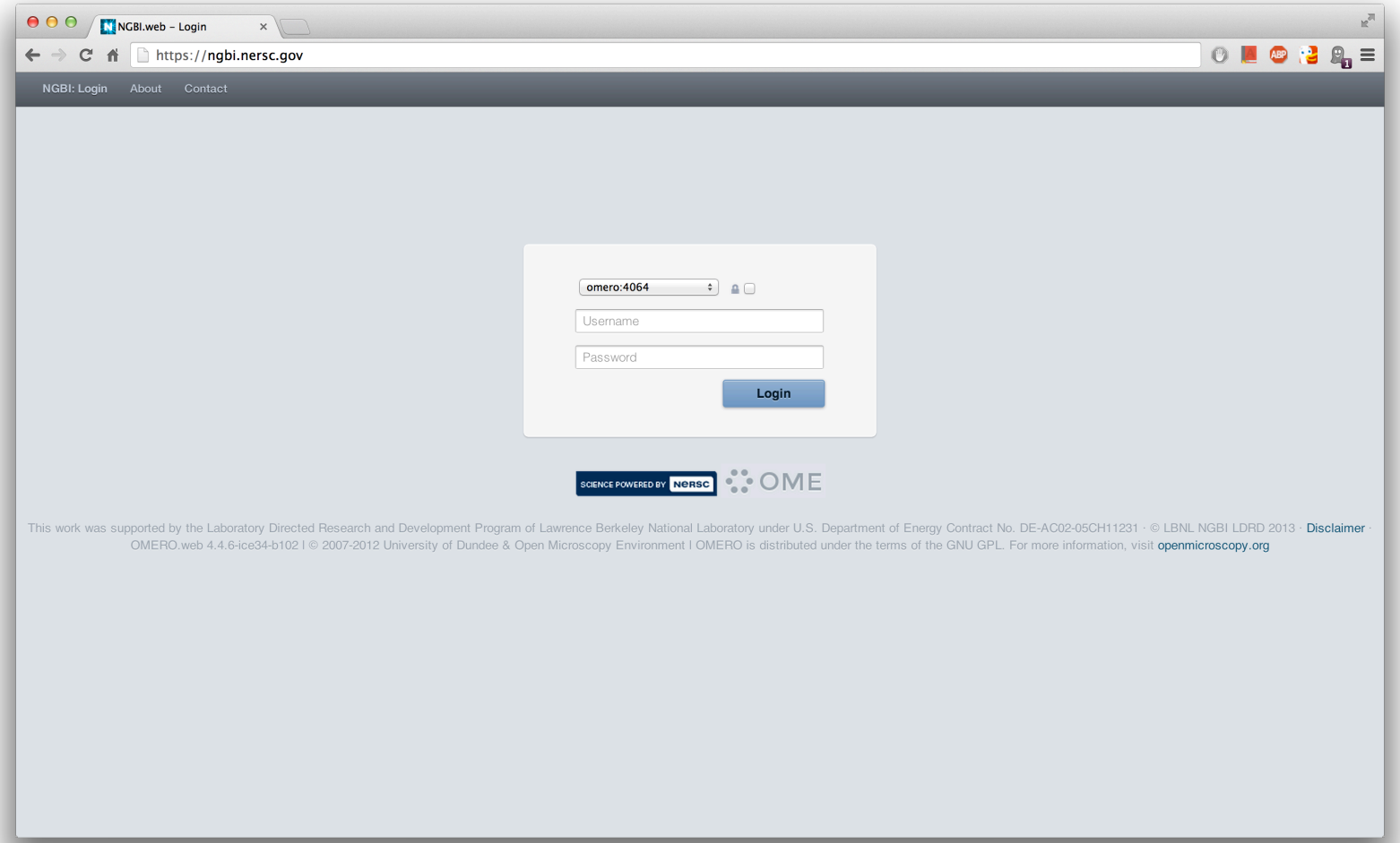


NERSC computer, storage and web systems support complex workflows that run thousands of simulations to screen materials ("Materials Genome"), proteins, structures and more; the results are shared with academics and industry through a web interface

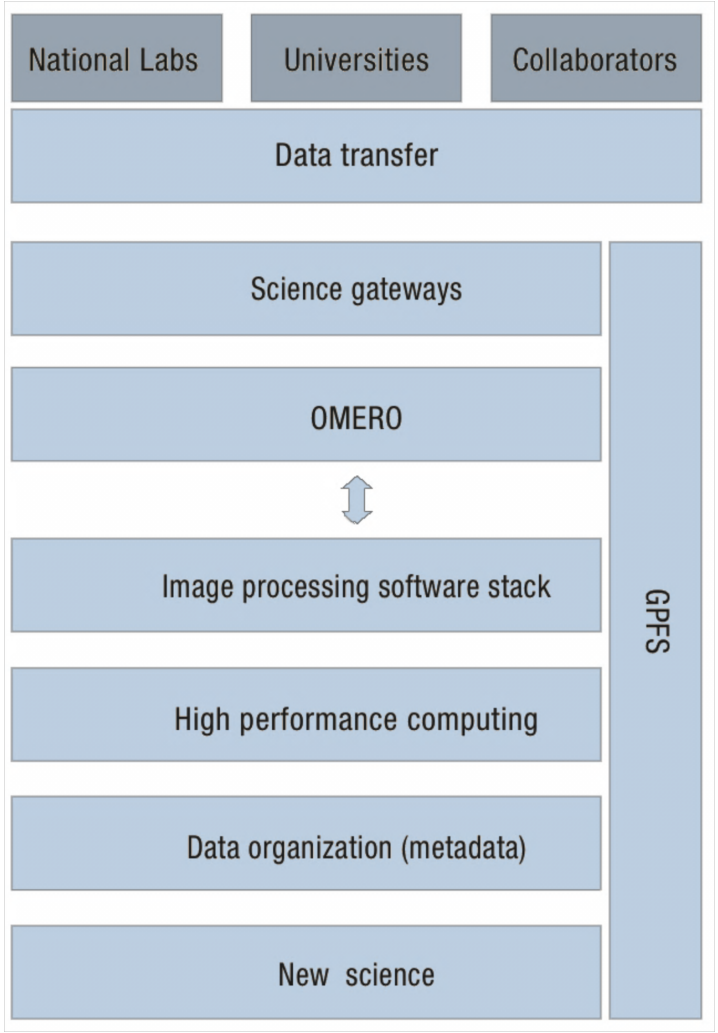
U.S. DEPARTMENT OF ENERGY | Office of Science

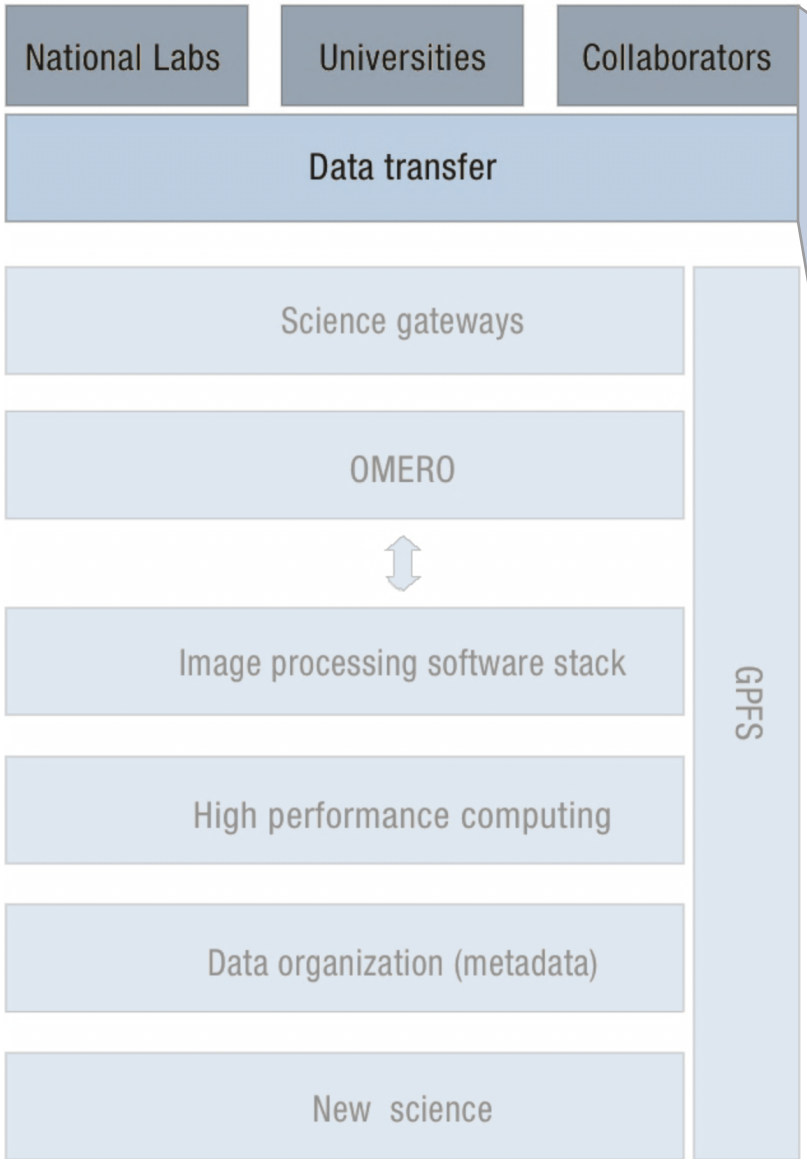
BERKELEY LAB
Lawrence Berkeley National Laboratory

Integrated tools for NGBI



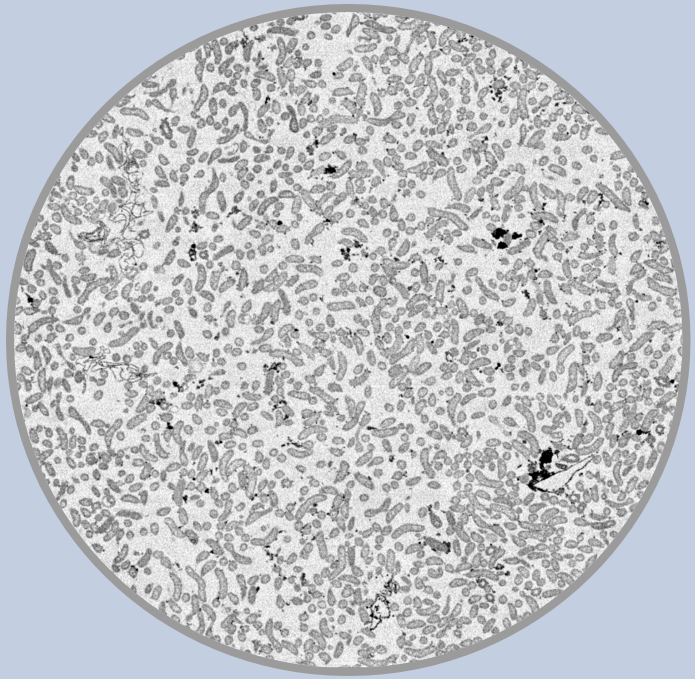
Architecture





NERSC

ACQUISITION



Raw image: Volume slice of an unstained sample of *Desulfovibrio vulgaris RCH1* showing extracellular metal deposits.

National Labs | Universities | Collaborators

Data transfer

Science gateways

OMERO

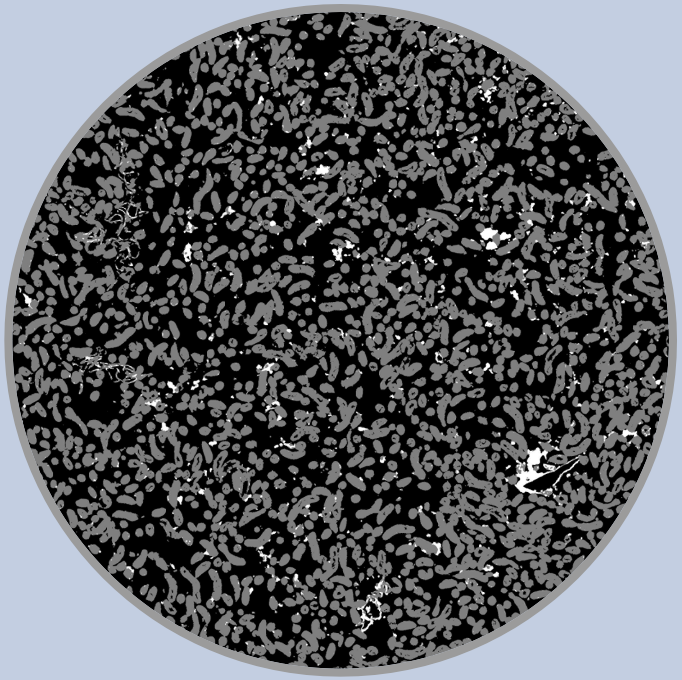
Image processing software stack

High performance computing

Data organization (metadata)

New science

SEGMENTATION



Segmented dataset showing a volume slice: background in black, bacteria in grey and metal deposits in white.

National Labs | Universities | Collaborators

Data transfer

Science gateways

OMERO

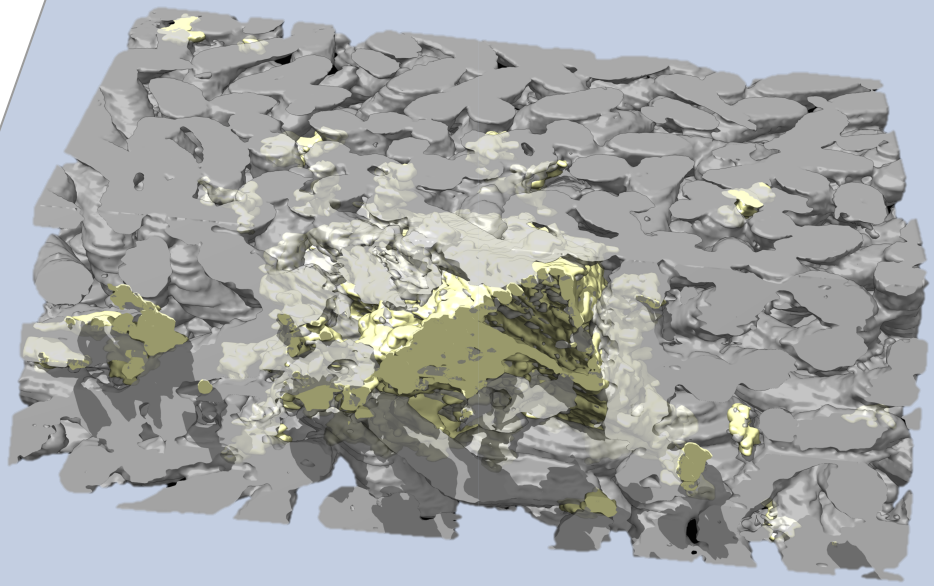
Image processing software stack

High performance computing

Data organization (metadata)

New science

3D VISUALIZATION



Visualization of ROI with bacteria shown in grey and extracellular metal deposits shown in yellow

National Labs | Universities | Collaborators

Data transfer

Science gateways

OMERO

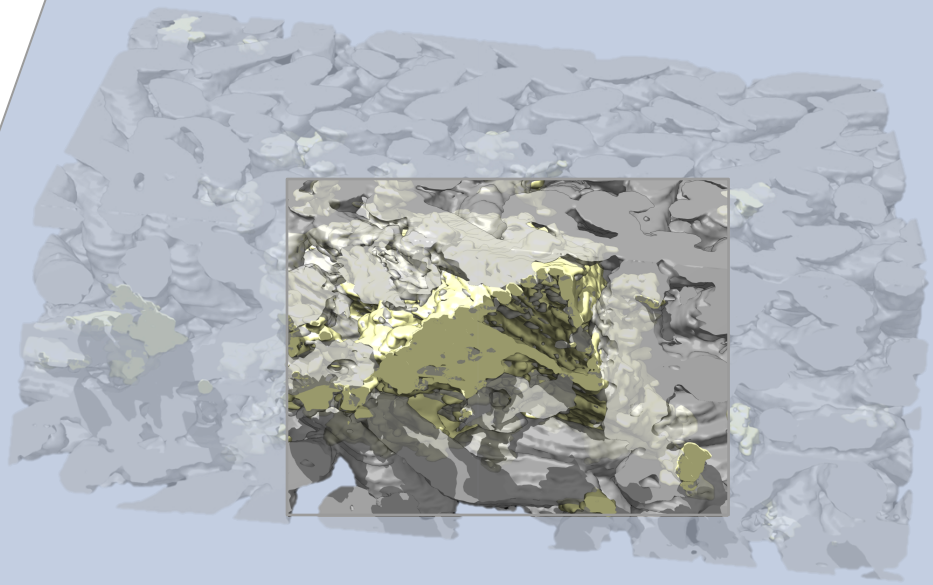
Image processing software stack

High performance computing

Data organization (metadata)

New science

VISUALIZATION



Visualization of ROI with bacteria shown in grey and extracellular metal deposits shown in yellow

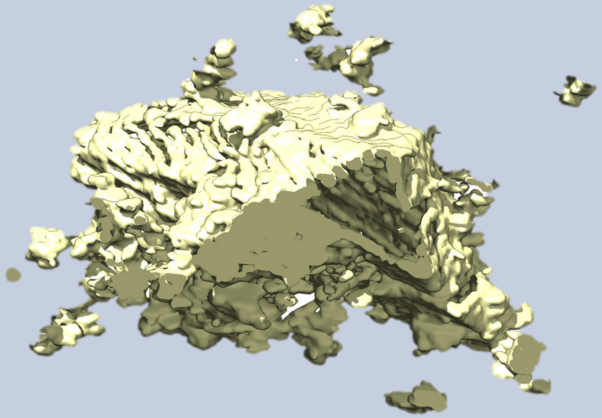
National Labs

Universities

Collaborators

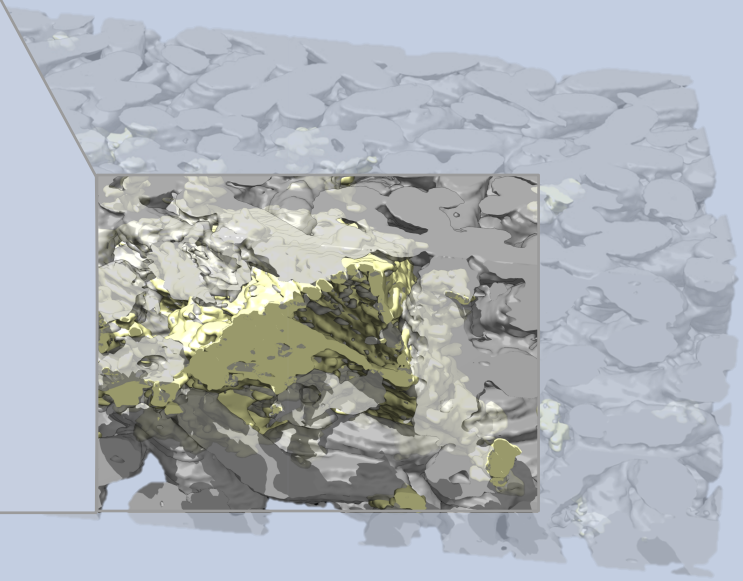
Data organization (metadata)

New science



Higher magnification of ROI sub-region showing the 3D organization of the metal deposits, with bacteria not being displayed

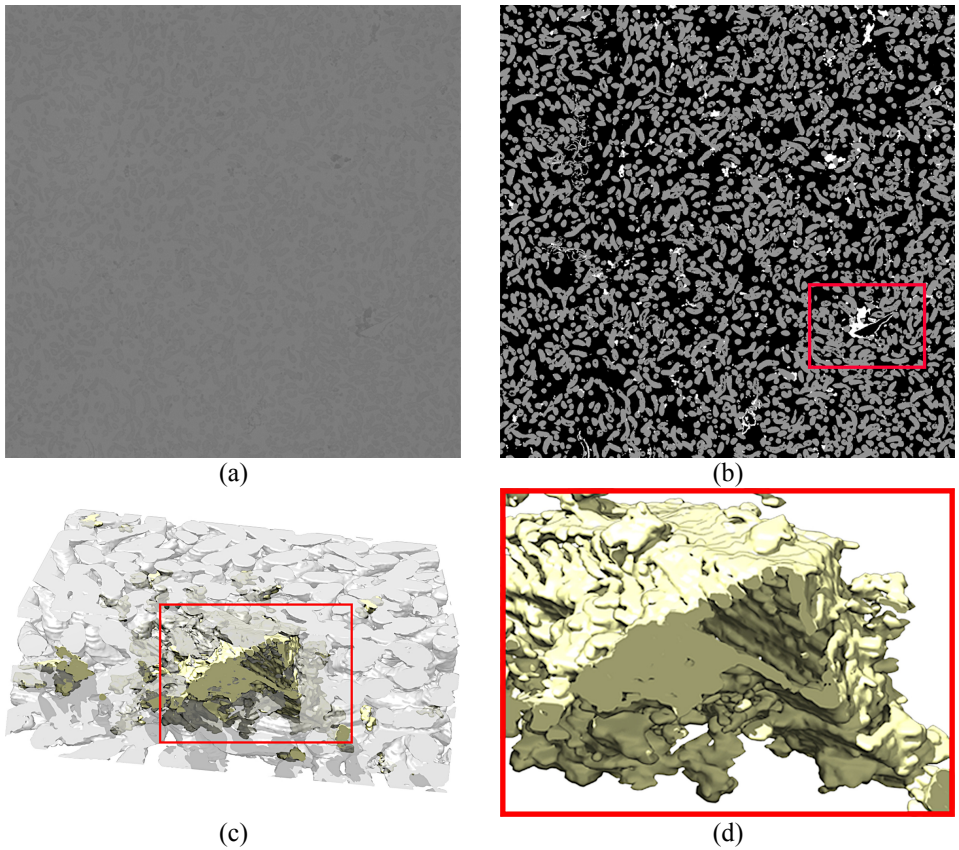
VISUALIZATION



Visualization of ROI with bacteria shown in grey and extracellular metal deposits shown in yellow

Results

- Novel insights in structural biology
- Unprecedented big data processing and handling (V: 40µm x 40µm x 100µm) and ~ 9GB/ dataset for CryoEM
- Understanding of complex system in high-resolution
- Multi-modal capabilities
- User-friendly web-based platform
- Metadata ready
- Robust automated approaches for large volumes.



Desulfovibrio vulgaris RCH1 (a) Raw image showing a slice of the volume (b) Segmented dataset showing a slice of the volume and ROI (c) 3D rendering of ROI showing bacteria and extracellular metal deposits (d) higher magnification of ROI sub-region of (c) showing the 3D organization of the metal deposits, with bacteria not being displayed

Acknowledgements



Shreyas Cholia *David Skinner* *Manfred Auer*

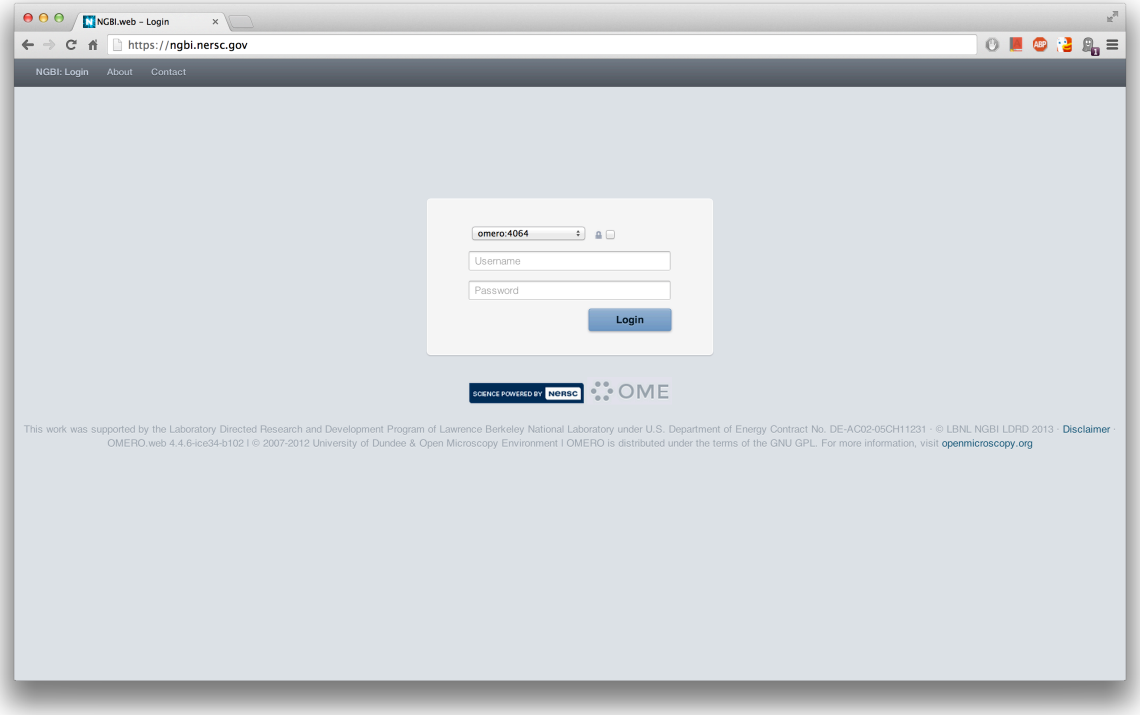
NERSC | LSD

This work was supported by the Laboratory Directed Research and Development Program of Lawrence Berkeley National Laboratory under U.S. Department of Energy Contract No. DE-AC02-05CH11231.

For more information:

http://ngbi.nersc.gov

Visit us



For more information:

http://ngbi.nersc.gov

JoaquinCorrea@lbl.gov

Big Data Bioimaging Architecture