Science Gateways at NERSC

Annette Greiner, NERSC Data and Analytics Services
What is a Science Gateway?

• Web-based scientific resource for collaboration
• A form of computer-supported cooperative work
• Can be shared data, shared tools, may include social networking
• “value-added interfaces to access these shared resources” (sciencegateways.org)
Harnessing the Power of the Web

• Traditional High Performance Computing assumes command-line expertise.

• The end-user scientist shouldn’t have to be an old-school unix geek.

• Take advantage of rich visualizations and advanced user interfaces.

• Facilitate better sharing and collaboration.
Some Examples
Some Examples

Interfaces to HPC resources
Some Examples

Data sharing across facilities/collaborations
Some Examples

Interactive tools

OpenMSI

Mass spectrometry imaging

NERSC

U.S. DEPARTMENT OF
Energy
Office of
Science

BERKELEY LAB

- 7 -
Some Examples

Rich visualizations and UIs
Three Levels of Collaboration

Self-Service

Initial Consulting

Immersive Engagement
User-Built Gateways

Maximizing data availability

Welcome to the Coherent X-ray Imaging Data Bank (CIXDB), a new database which offers scientists from around the world a unique opportunity to access data from Coherent X-ray Imaging (CXI) experiments.

- New light sources and detectors have enabled novel experiments producing terabytes of data per day.
- It's the dawn of the data deluge era for coherent X-ray imaging.
- To best make use of all these data it is necessary to make them accessible.

Accessibility is crucial not only to make efficient use of experimental facilities, but also to improve the reproducibility of results and enable new research based on previous experiments.

CIXDB is dedicated to further the goal of making data from Coherent X-ray Imaging (CXI) experiments available to all, as well as archiving it. The website also serves as the reference for the CXI file format, in which most of the experimental data on the database is stored in.

We must all accept that science is data and data are science, and thus provide for [...] much improved data curat...
Building Your Own Gateway

What level of service do you need? (We are 8x5)

Do you need to work with a programmer? (Domain Expertise + CS expertise)

What kind of data do you have? (HDF5, Database, Files)

Do you have long running computations? (> web time)

Would you benefit from an API? (Making data available in a structured format)
Building Your Own Gateway

The quick way:
World-readable files in /project/projectdirs/myproject/www

/project/projectdirs/myproject/www/index.html

http://portal.nersc.gov/project/myproject/index.html

<!DOCTYPE html>
<html lang="en">
<head>
  <meta charset="utf-8">
  <title>title</title>
</head>
<body>
  <h1>Hello, world!</h1>
</body>
</html>
Django and Flask (Python) are popular web frameworks

Python has an extensive set of scientific libraries like NumPy, SciPy, H5PY

See also: PHP, web2py, etc.

PyDAP for serving HDF5 data

Jupyter Notebook, R-Studio for Interactive Computing

Globus, NEWT
Web APIs

We’ve found that scientists often want programmatic access to data
e.g. Materials Project: Give me property X for all materials with Li and O so that I can pass it through my own codes
Lesson – make your data available through an API and people will start to do new and innovative things
Web API Example

GET https://www.materialsproject.org/rest/v1/materials/Fe2O3/vasp/energy

{
    "created_at": "2013-03-17T09:14:58.158081",
    "valid_response": true,
    "version": {
        "pymatgen": "2.5.4",
        "db": "2013.02.25",
        "rest": "1.0"
    },
    "response": [{
        "energy": -132.33005625,
        "material_id": 542309
    }, {
        "energy": -66.62512425,
        "material_id": 24972
    }],
    "copyright": "Copyright 2012, The Materials Project"
}
NEWT

REST API for HPC resources
NERSC over HTTP
Jobs, Files, Commands, Status, Accounting, etc.
Makes it very easy to build web applications that can interface with NERSC
https://newt.nersc.gov
NEWT Example

GET https://newt.nersc.gov/newt/status/

[{
    "status": "up",
    "system": "cori"
}, {
    "status": "up",
    "system": "edison"
}, {
    "status": "up",
    "system": "pdsf"
}, {
    "status": "up",
    "system": "genepool"
}, {
    "status": "up",
    "system": "archive"
}]}
Dogfooding
What Next?

Read the Docs:
http://www.nersc.gov/users/science-gateways/
Contact the science gateways team

Shreyas Cholia
Access, Management

Rollin Thomas
Astrophysics, Cosmology
Analytics, Access

Annette Greiner
Biomedical Sciences
Access, Visualization

consult@nersc.gov