Collaborative Infrastructure to Integrate Local and Cloud Resources and Ease Scientific Collaboration

K.-S. Kuo[1,2], R. Ramachandran[3], M. Maskey[4], C. Lynnes[2], N. Niamsuwan[5]

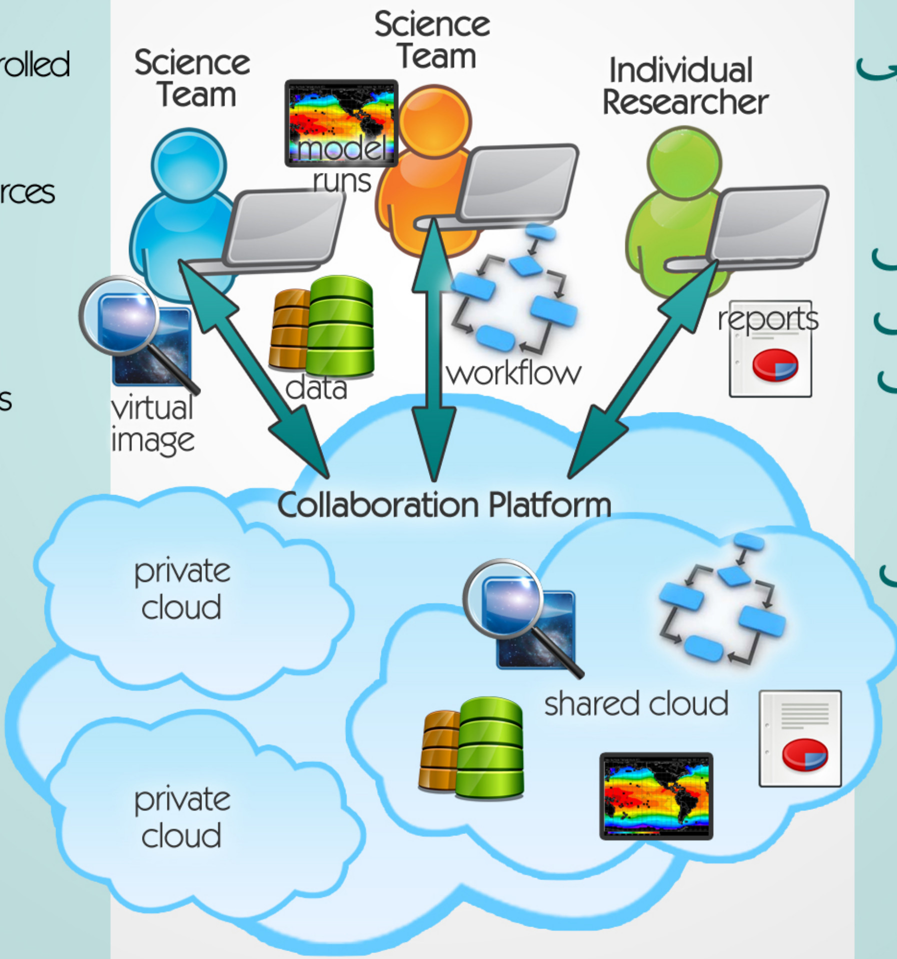
[1]Bayesics, LLC; [2]NASA Goddard Space Flight Center; [3]NASA Marshall Space Flight Center; [4]University of Alabama-Huntsville; [5]NASA Jet Propulsion Laboratory

Collaborative Workbench (CWB)
to Accelerate Science Algorithm Development

Sharing Knowledge is at the heart of science, yet it is challenging for researchers to effectively share information and tools

Goals

- An architecture for scalable, controlled collaboration
- Selective sharing of science resources
  - among individuals
  - within science teams
  - with the entire science community.
- Software that fits how researchers currently do scientific analysis to promote adoption



Benefits

- Accelerate science algorithm development by distributed science teams
- Reduce redundancy
- Improve productivity
- Securely share all science artifacts (data, information, workflow, virtual machines)
- Generalizable to support collaborative science algorithm development for other mission and model enterprises

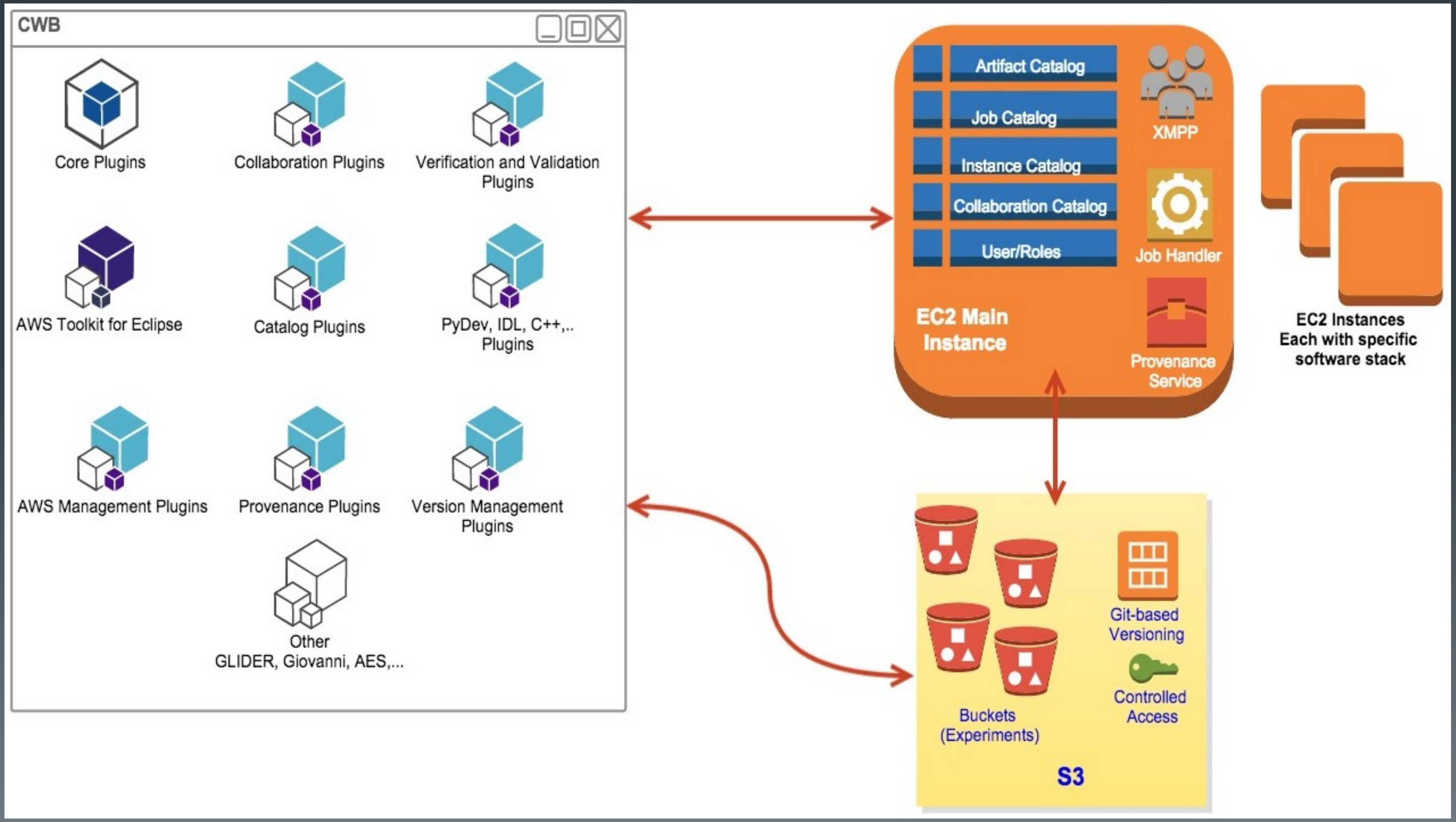
Significance of Collaborative Workbench (CWB)

- Earth Science is a science for “system of systems.”
  - Earth scientists are required to collaborate outside of their traditional focus areas.
- Collaborations are increasingly geographically distributed across nations and time zones.
  - GPM (and TRMM) is a prime example.
- We need to accommodate various modes of collaboration, in order for it to be effective.
- Collaboration needs to be made easy so scientists can focus on science, rather than learning collaboration tools.

Outline

- CWB core components
- Logging on to CWB infrastructure.
- Creating experiments and workflows.
- Collaborating asynchronously.
- Collaborating synchronously.
- AES Integration
- NEOS^3 Integration – Creating NEOS^3 jobs
- NEOS^3 Integration – Obtaining NEOS^3 input and output

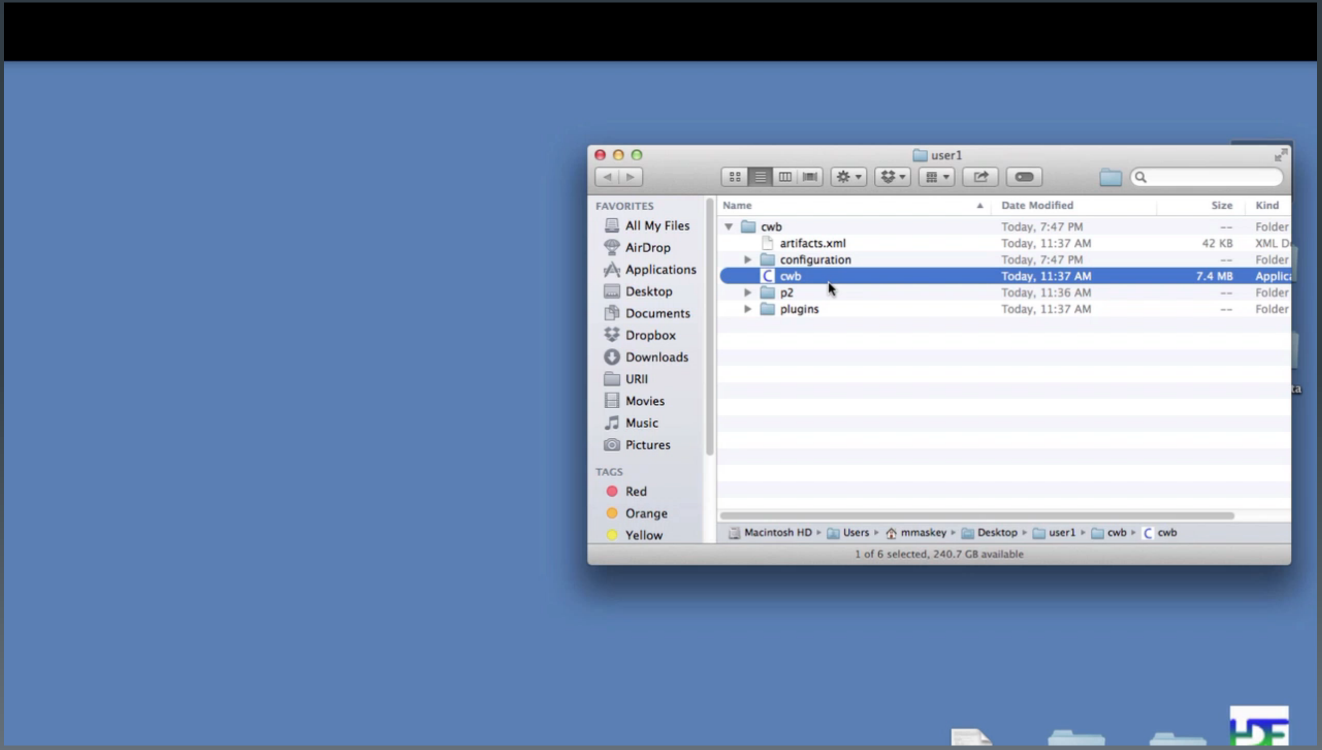
CWB Core Components



Logging On

- CWB seamlessly integrates your local compute resources with those in Cloud.
  - A CWB server keeps track of collaboration group membership and Cloud resources, so an account (and logging on) is required.
- An ***experiment*** is composed of one or more workflows.
- A ***workflow*** is a sequence of one or more programs (applications), which are executed in order to accomplish a specific task.
- Three (3) “workspaces”
  - ***Shared Experiments***: Experiments shared by others
  - ***My Experiments***: My personal workspace
  - ***Other Experiments***: Experiments created by others but no workflow has been shared.

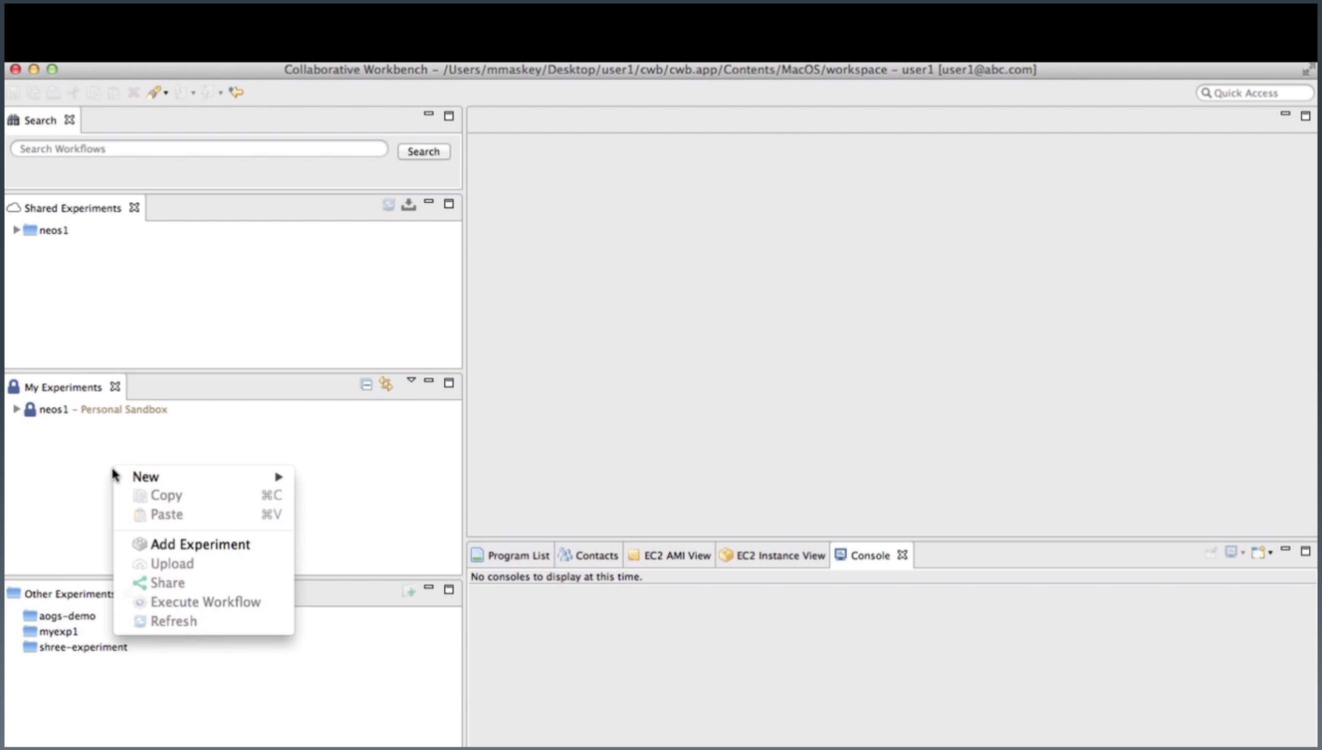
Logging On


user1
FAVORITES
All My Files
AirDrop
Applications
Desktop
Documents
Dropbox
Downloads
URII
Movies
Music
Pictures
TAGS
Red
Orange
Yellow
Name
Date Modified
Size
Kind
cwb
Today, 7:47 PM
--
Folder
artifacts.xml
Today, 11:37 AM
42 KB
configuration
Today, 7:47 PM
--
Folder
cwb
Today, 11:37 AM
7.4 MB
p2
Today, 11:36 AM
--
Folder
plugins
Today, 11:37 AM
--
Folder
Macintosh HD ▸ Users ▸ mmaskey ▸ Desktop ▸ user1 ▸ cwb ▸ cwb
1 of 6 selected, 240.7 GB available

Creating Experiments and Workflows

- Code and data can be dragged and dropped into CWB personal workspace.
- Code may be executed *either* locally *or* in Cloud.
- Users can choose different Cloud instances (different software configurations, e.g. different IDL versions).

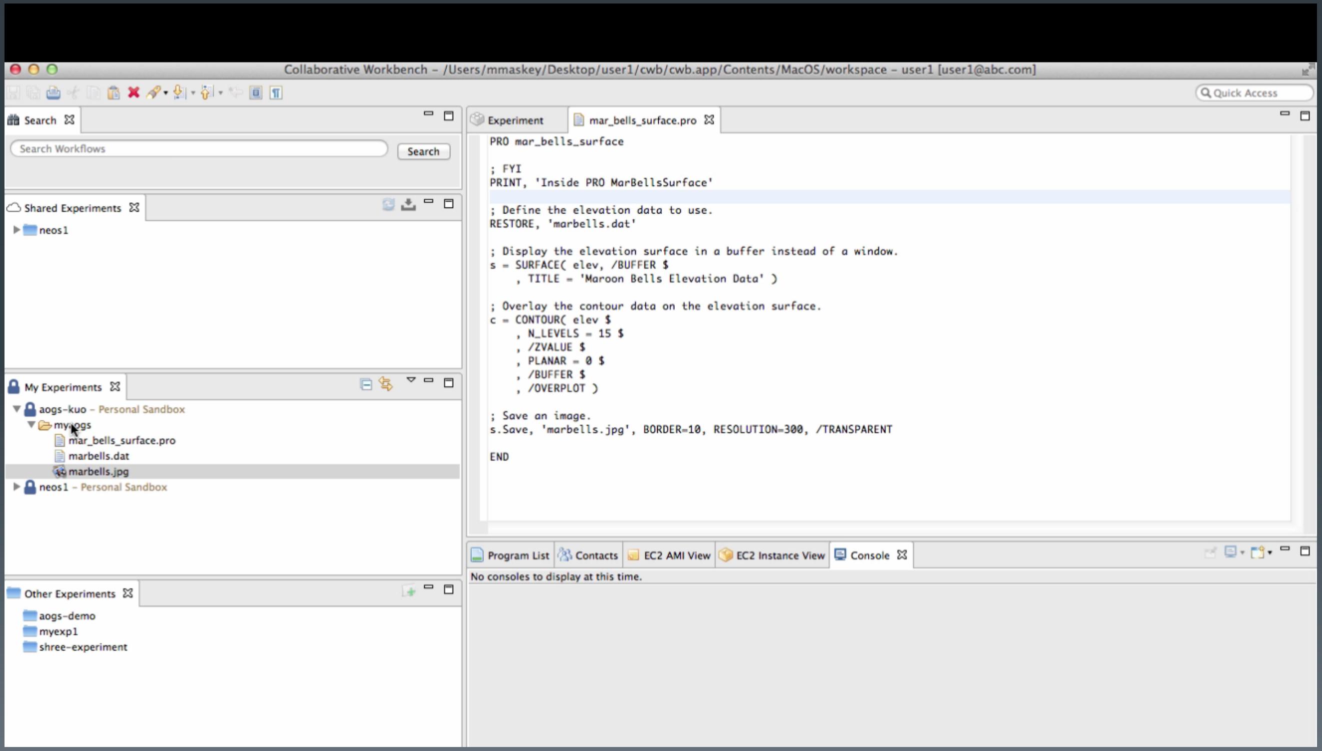
Creating Experiments and Workflows



Collaborating Asynchronously

- Members of the same group can execute shared experiments and workflows at their own convenience.

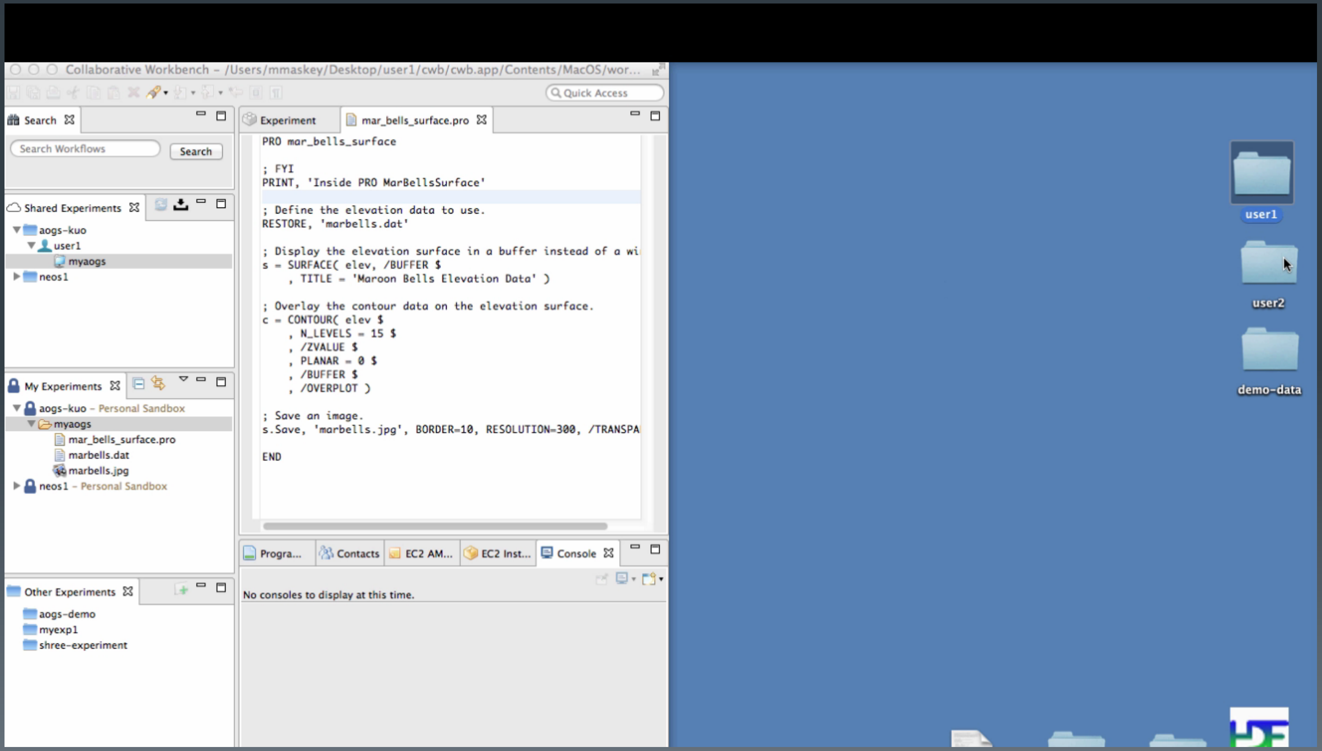
Collaborating Asynchronously



Collaborating Synchronously

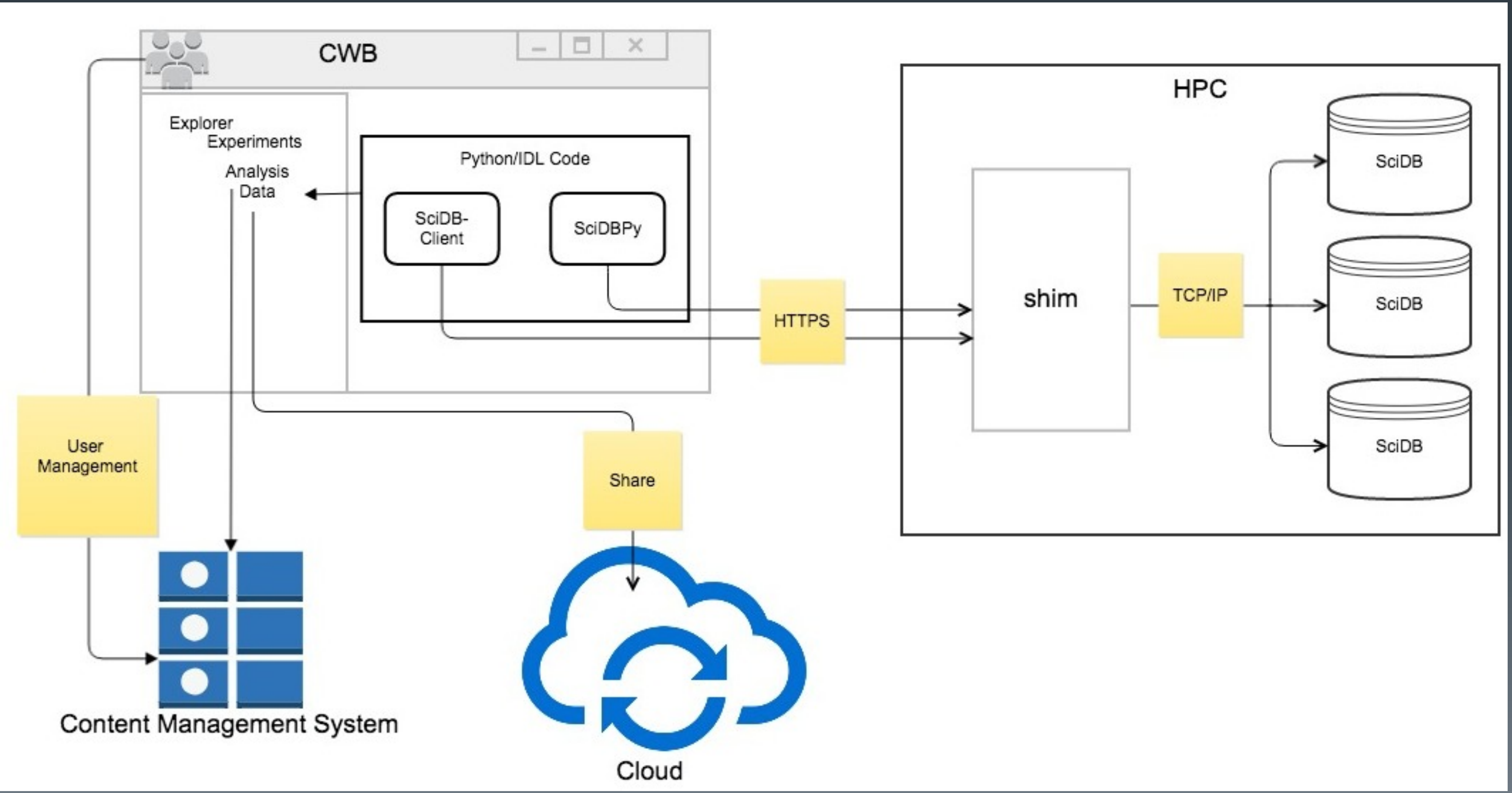
- Often asynchronous collaboration is inadequate and synchronous collaboration is required.
- Users need to sign into CWB and connect with an XMPP server (provided by CWB infrastructure).
  - XMPP: eXtensive Messaging and Presence Protocol
- Once connected, multiple users in the same group can chat and share in real time.
- Two (2) members at a time can engage in ***collaborative programming***.
  - In the upcoming clip, user1 is on the left and user2 on the right.

Collaborating Synchronously



14 November 2014
HPC Symposium, Trieste, Italy

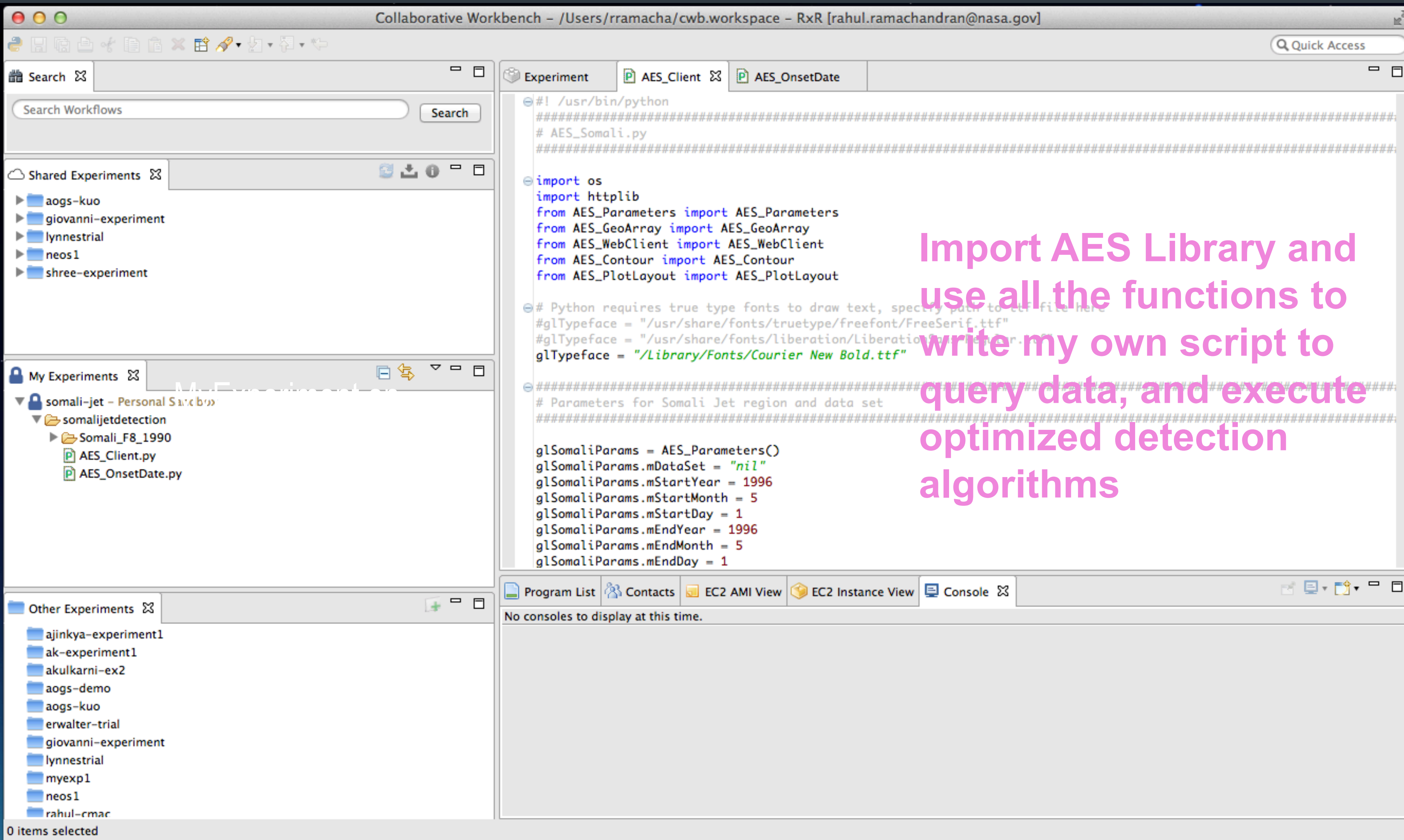
CWB-AES Integration



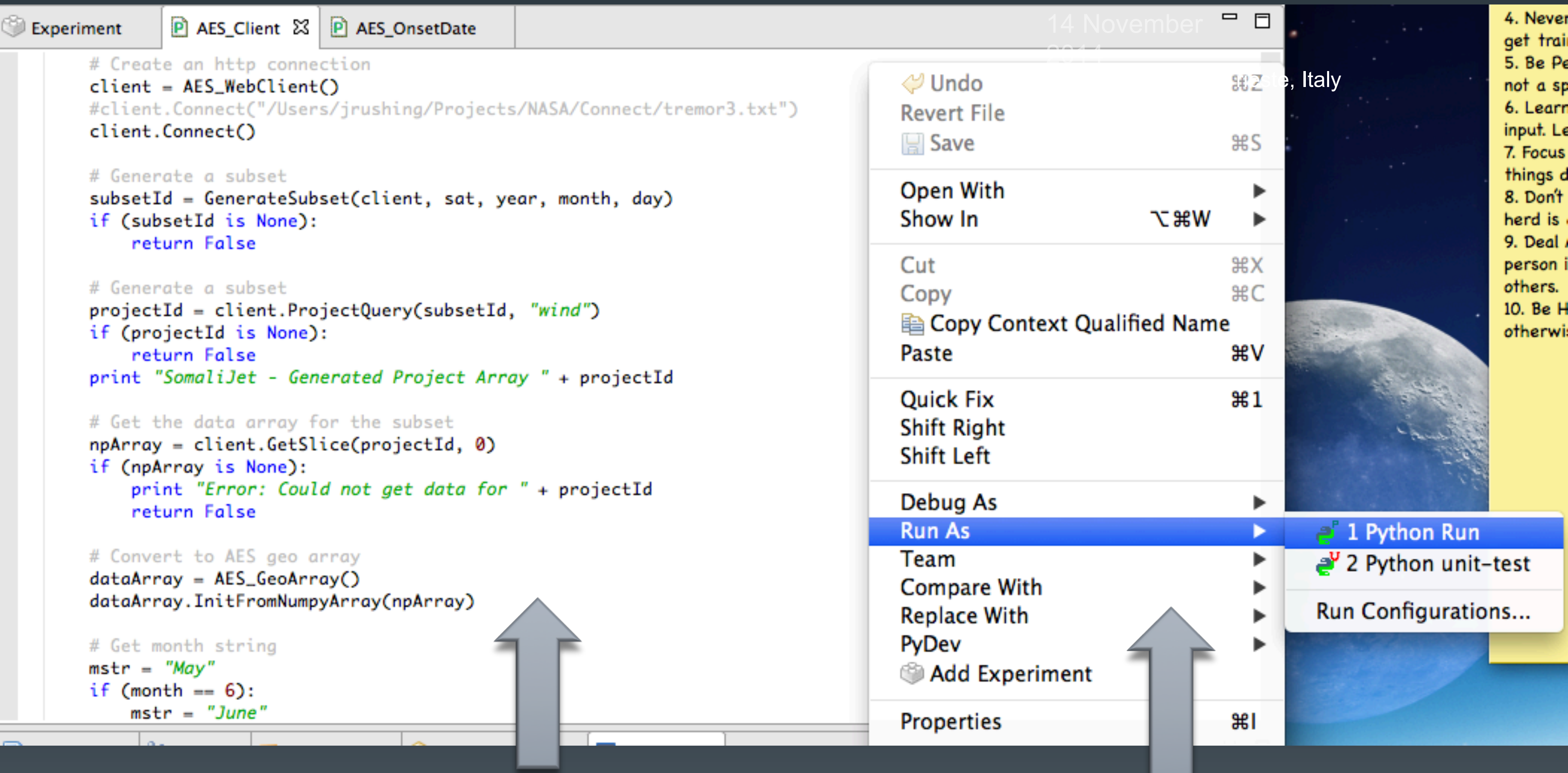
Starting CWB

Authentication and authorization handled by CWB

Point to the appropriate HPC End Point

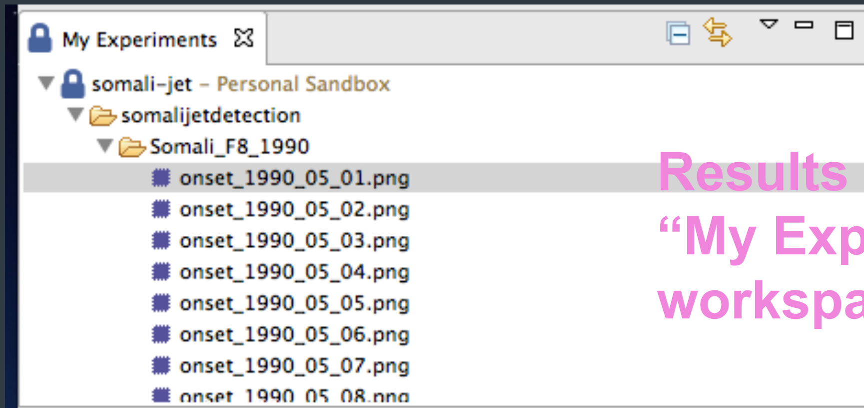


Import AES Library and use all the functions to write my own script to query data, and execute optimized detection algorithms

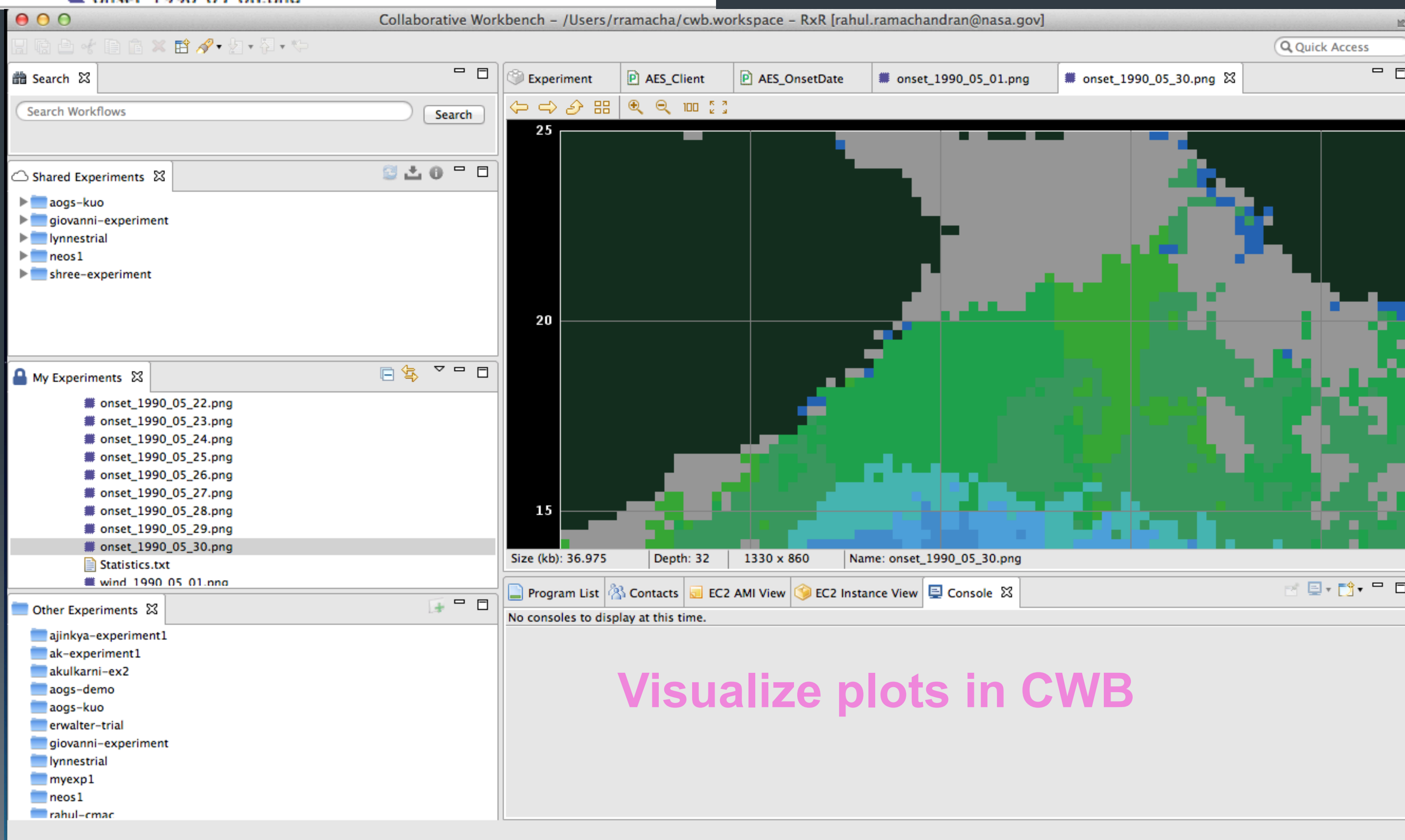


- **Execute optimized data parallel queries on a remote HPC**
- **Integrate local data processing and analysis seamlessly**

Run my script from CWB

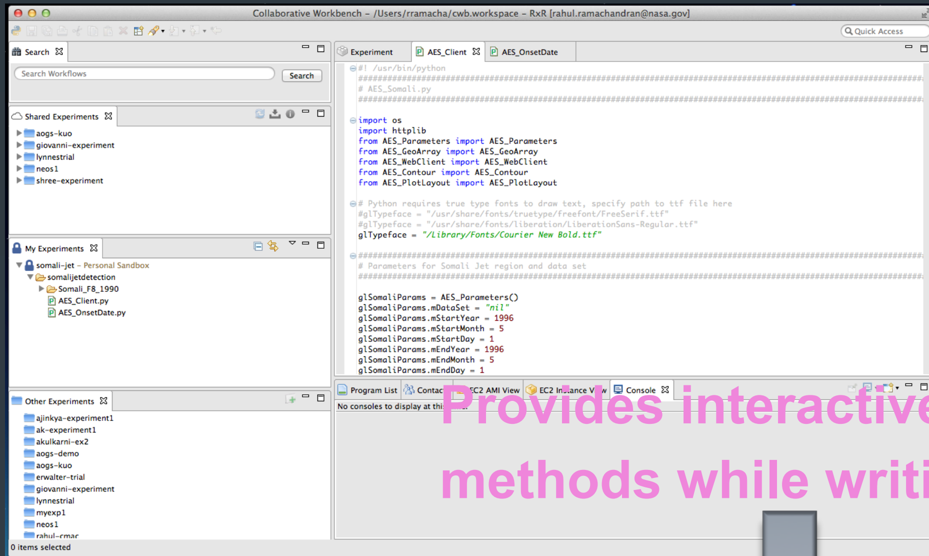


Results saved in "My Experiment" workspace

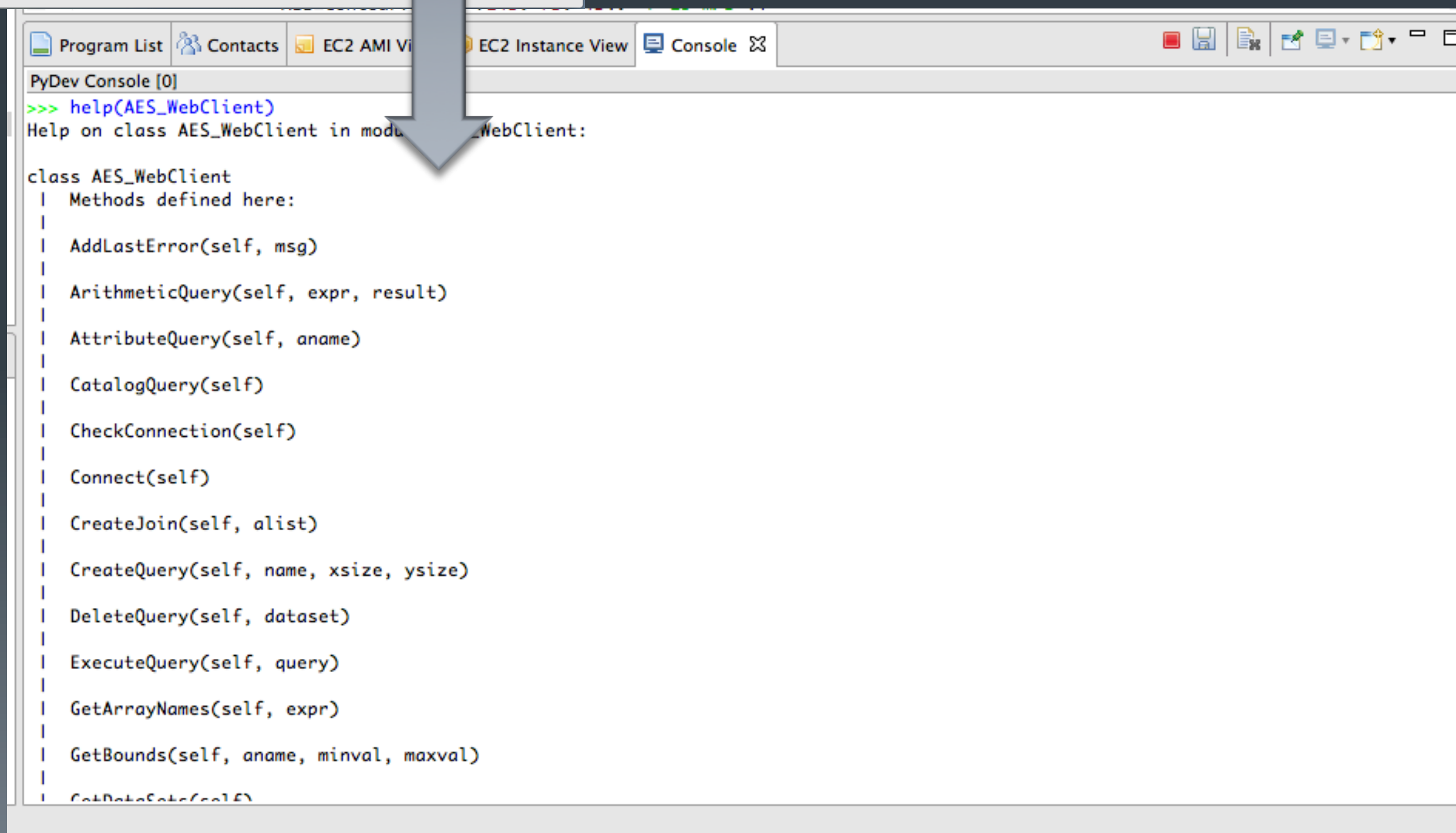


Visualize plots in CWB

14 November 2014
HPC Symposium, Trieste, Italy



Provides interactive Python shell to try methods while writing the script

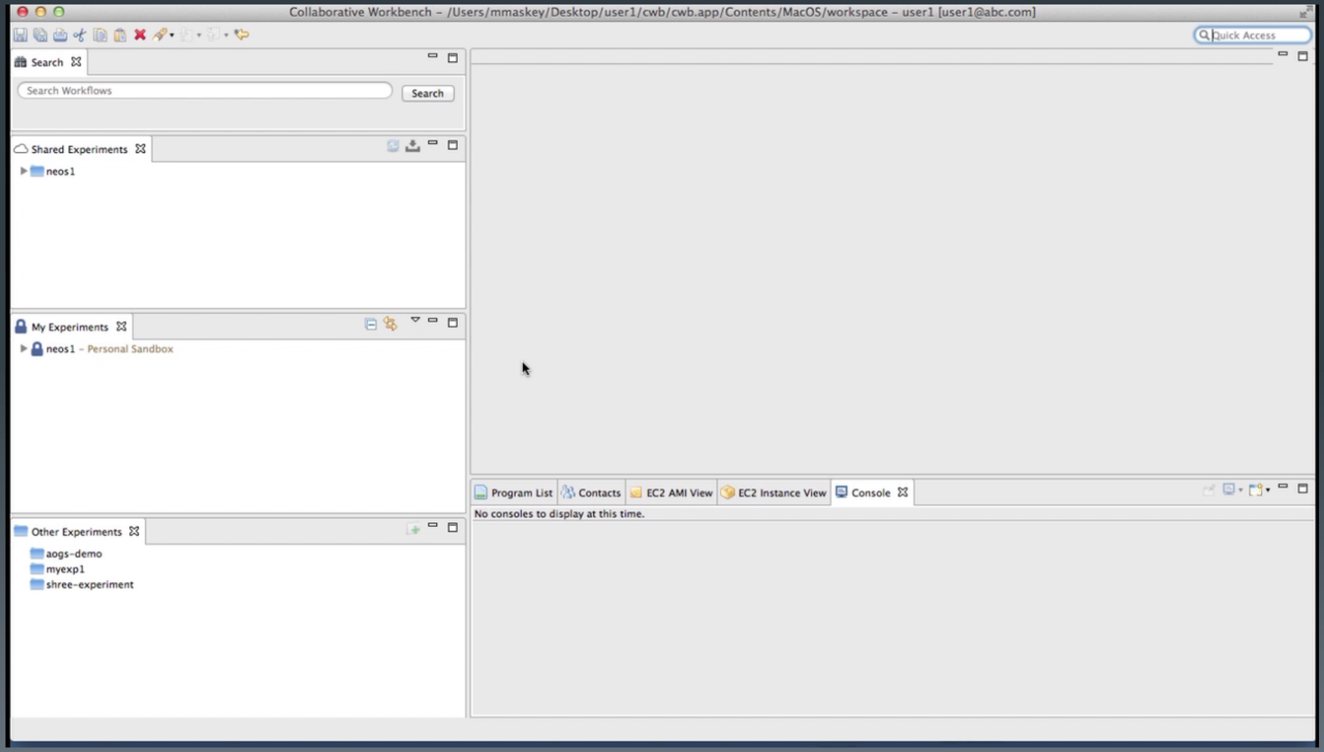


NEOS^3 Integration
Creating NEOS^3 Jobs

- NEOS^3: NASA Earth Observing System Simulation Suite
  - A project PI'ed by ***Simone Tanelli*** of JPL and funded by NASA Advanced Information System Technology (AIST) program.
  - Web-based integrated simulator for Earth remote sensing applications.
  - Equipped with start-of-the-art modules to enable the realistic simulation of satellite observables.
  - Providing an advanced, sophisticated, and user-friendly simulator package to be used by both ***scientists*** for research-oriented applications and by ***system engineers*** for an instrument design purpose.
  - Accessible via a web interface and capable of distributing computationally intensive tasks to remote servers such as those at the NASA Advanced Supercomputing (NAS) Division.

14 November 2014
HPC Symposium, Trieste, Italy

NEOS^3 Integration
Creating NEOS^3 Jobs

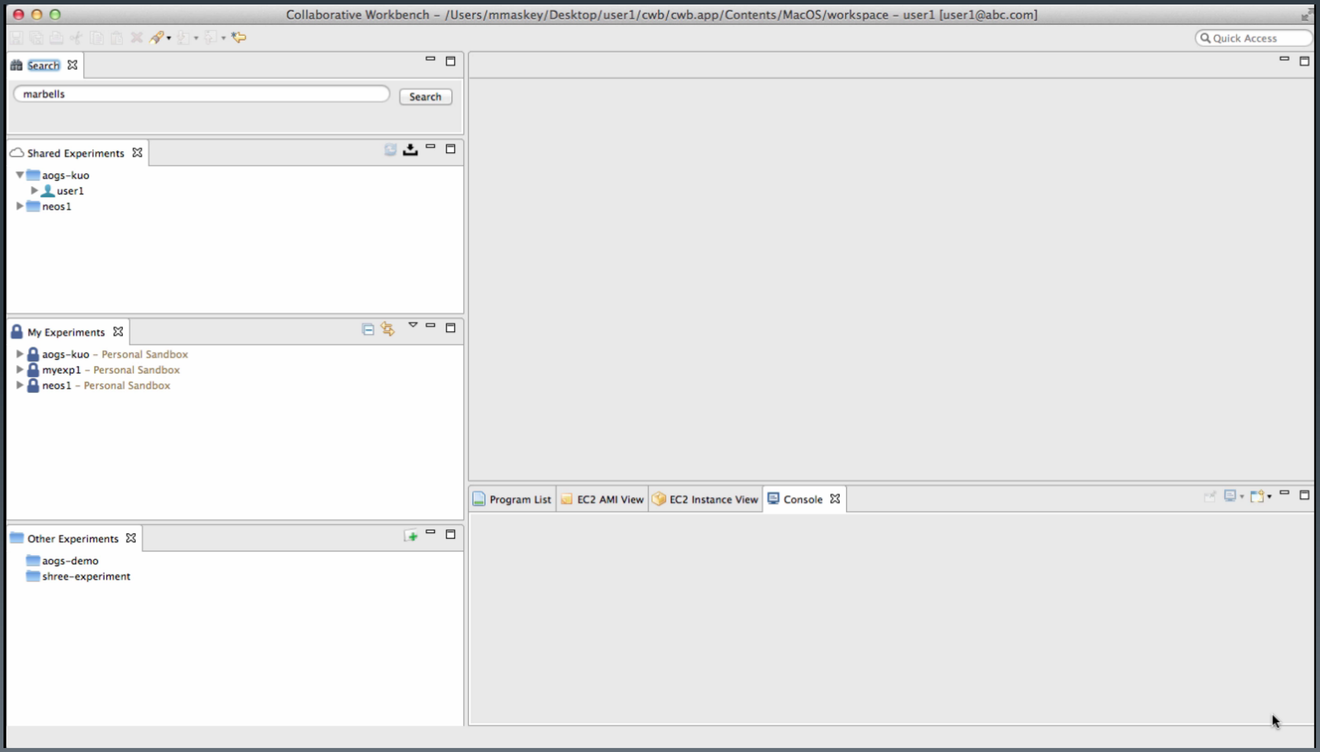


NEOS^3 Integration
Obtaining NEOS^3 Input/Output

- A NEOS^3 job may take hours.
- NEOS^3 module polls the server using Job ID until job is complete.
- Job output can then be retrieved using a function provided in NEOS^3 module.

14 November 2014
HPC Symposium, Trieste, Italy

NEOS³ Integration
Obtaining NEOS³ Input/Output



Epilogue

- CWB implements automatic versioning (in the background) by integrating GitHub.
  - Similar to GoogleDoc, you can always go back to a previous version.
- CWB has integrated
  - Automated Event Service.
  - NASA Earth Observing System Simulator Suite, NEOS^3.
  - NASA GES DISC Giovanni service.
- CWB is currently in alpha test.

Acknowledgement

- NASA Computational Modeling Algorithms and Cyber-infrastructure (CMAC) program
- NASA High-End Computing (HEC) program
- NASA Earth Science Technology Office (ESTO)

14 November 2014
HPC Symposium, Trieste, Italy

Adding Workflow(s) in Experiment(s) of Others

14 November 2014
HPC Symposium, Trieste, Italy

Adding Workflow(s) in Experiment(s) of Others