

# Florida International University



PARTNERSHIP FOR INTERNATIONAL  
RESEARCH AND EDUCATION

TERENA June 3<sup>rd</sup>, 2013

Julio Ibarra, PhD.

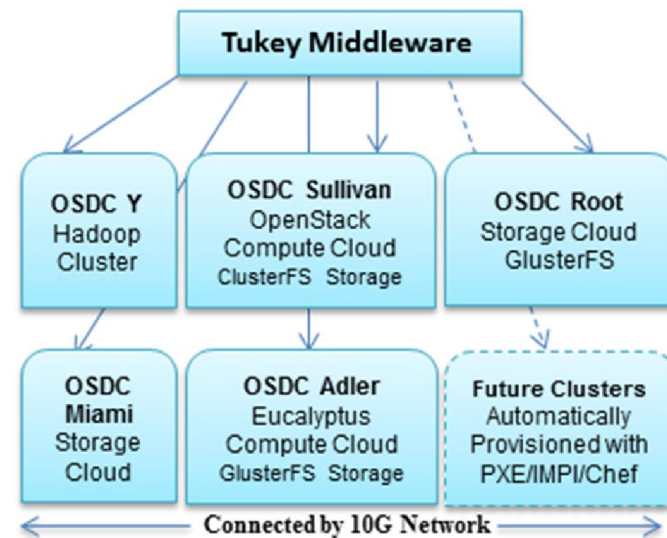
Assistant Vice President of Technology Augmented Research (CIARA)



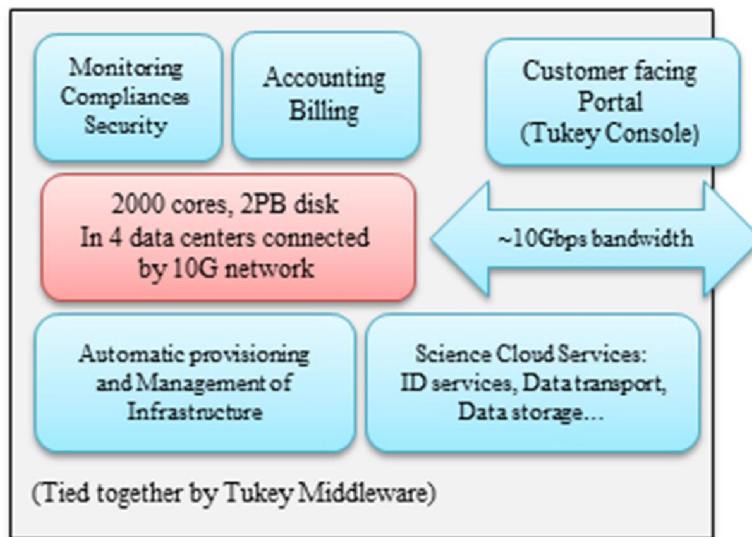
# Overview of OSDC

The Open Science Data Cloud (OSDC) is an open-source, cloud-based infrastructure that allows scientists to manage, analyze, integrate and share medium to large size scientific datasets.

The OSDC is a distributed facility connected by 10G or greater networks so high speed transport protocols are important for enabling users to import/export data and to move data around flexibly in their analysis processes.



# Overview of OSDC



The OSDC user services include the ability to provision virtual machines, access usage and billing information, share files, and access to a key service and public datasets. All of the OSDC user services are tied together through a customizable web application "Tukey" and middleware that enables uniform access to the cloud services by the "Tukey Console" web application.

# From bare metal to a compute or storage cloud

Using Chef starts with one [PXE boot server](#), a [Chef server](#), and a set of servers with [IPMI](#) configured. IPMI is triggered to boot the servers, which then pull a start-up image and boot options from the [PXE boot server](#). The options define the location of a preseed file which the start-up sequence loads and uses as a guide to automate the installation of the generic Ubuntu Server.

## Initialization

The installation uses the PXE boot server or a preconfigured proxy to install Ubuntu Server directly from the repositories to avoid any updates being required post-install. Then the installer runs a script specified at the end of the preseed file which sets up networking on the freshly installed system and adds another script to be run at boot.

## Installation

Upon rebooting, the next script double-checks the IPMI configuration, finishes partitioning the disk and sets up additional RAID as necessary, before downloading and installing the Chef client.

## After-Install

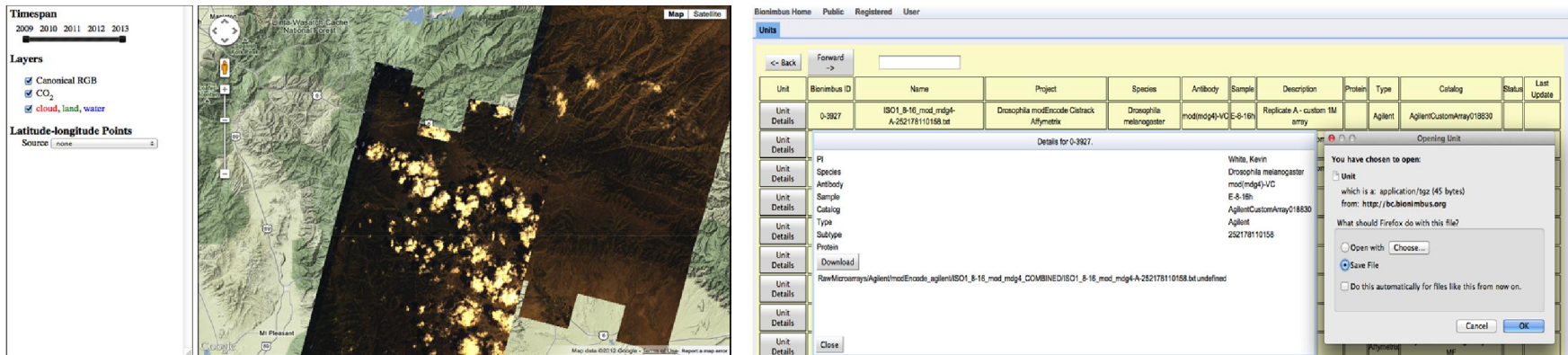
The Chef client then checks in with the Chef server and runs the “recipes” listed for either a management node or a compute node. As a last step, a final clean up script runs to deliver us a fully functional OpenStack rack.

## Check

[UDT](#) | [UDR](#) | [Debian Linux 5.0](#) | [Ubuntu 12.04](#) | [Open Stack](#) | [Eucalyptus](#) | [Sector/Sphere](#) | [Hadoop/MapReduce](#) | [MySQL](#) | [PostgreSQL](#)



# OSDC-Projects



*Project Matsu* is a collaboration between NASA and the Open Cloud Consortium to develop open source technology for cloud-based processing of satellite imagery to support the earth sciences.

*Bionimbus* is a cloud-based infrastructure for managing, analyzing, archiving, and sharing large genomic databases. Bionimbus is used by a number of projects, including *modENCODE*, *ENCODE*, and the *T2D-Genes* consortia.

Computer Science | Bioinformatics | Electrical Engineering | Medicine | Astronomy | Earth Science | Physics | Genetics | Biology | Chemistry





# OSDC-PIRE Int'l Partners



The Open Science Data Cloud PIRE project provides international research and education experiences through training and study at universities and research institutes around the world with leading scientists in the field of computing. Our goal is to increase the participants with expertise in managing and analyzing data. The creation of a strong cadre of students with a global perspective on scientific data management will support research endeavors in many other areas vital to the US. International Collaboration.



[Beijing Institute of Genomics \(BIG\), Chinese Academy of Sciences](#) | [National Institute of Advanced Industrial Science and Technology \(AIST\), Japan](#) | [Universidade Federal Fluminense, Rio de Janeiro \(UFF\)](#) | [University of Amsterdam](#) | [University of Edinburgh](#) | [USP - Universidade de São Paulo](#)



# Questions?

- Visit the Open Science Data Cloud website for more information on the [application](#), and for a list of important dates <http://pire.opensciencedatacloud.org/>
- Contact Robert Grossman at [robert.grossman@uchicago.edu](mailto:robert.grossman@uchicago.edu) with any questions or comments you may have.
- Contact Heidi Alvarez at [heidi@fiu.edu](mailto:heidi@fiu.edu) with any questions or comments you may have.
- Contact us at [pire@opensciencedatacloud.org](mailto:pire@opensciencedatacloud.org) if you have questions regarding the application process, important dates, or any other project specific inquiries.

