



---

Managed by Fermi Research Alliance, LLC for the U.S. Department of Energy Office of Science

---

# PDACS – A Portal for Data Analysis Services for Cosmological Simulations

Ryan Chard, **Saba Sehrish**, Alex Rodriguez, Ravi Madduri,  
Thomas D. Uram, Marc Paterno, Katrin Heitmann, Shreyas Cholia,  
Jim Kowalkowski, Salman Habib

*GCE: The 9th Gateway Computing Environments Workshop*

[SC14: The International Conference on High Performance Computing, Networking, Storage and Analysis](#)



# Introduction

---

- Accessing and analyzing data from cosmological simulations is a major challenge
  - large size data sets (100s of TBs)
  - diversity of the associated large-scale analysis tasks.
- PDACS is a web-based workflow service and scientific analysis platform for cosmology. Major contributions are:
  - Providing a scalable platform for a large set of cosmological analytical tools and facilitate parallel job submission over HPC infrastructure at NERSC and ANL.
  - Repurposing Galaxy as a cosmology-specific workflow service and research gateway.
- Galaxy is “an open, web-based platform for data intensive biomedical research”.

# Features of PDACS

---

- New Cosmology Tools
  - Provides a set of frequently used cosmology tools
  - Provides means to reuse tools and contribute new tools
- New Data types for metadata propagation
  - SQLite-based data type
- Data Access
- Job Submission
  - NERSC (NEWT API), ANL (Shibboleth)
- Plotting
  - JavaScript based GNU Plot
  - Shiny: web application framework for R
    - User interface for multiple dataset selection, plot controls, column selection, etc.

# Workflow Screenshots

**Galaxy** Analyze Data **Workflow** Shared Data Visualization Admin Help User Using 22.4 GB

Tools Workflow Canvas | imported: sc14-6

search tools

**Globus**  
**Get Data**  
**Halos - Simulation Data Analysis Tools**

- **Halo Finder** FOF/SO Halo Finder
- **c-M relation** Measure c-M relation from the SO halo dataset
- **FOF Mass Function** Measure FOF mass function on FOF halo dataset
- **SO Mass Function** Measure SO mass function on SO halo dataset

**Halos - Predictors**  
**2-point Functions - Simulation Data Analysis Tools**  
**2-point Functions - Predictors**  
**Conversion Tools**  
**Graph/Display Data**

**Workflow control**  
**Inputs**

**Select Snapshot** x  
 SelectedSnapshot (dbm)

**Halo Finder** x  
 Input Snapshot  
 FOFProperties (haloFinder\_binary)  
 SOProperties (haloFinder\_binary)  
 HaloParticles (haloFinder\_binary)  
 ParticleHaloTagFile (haloFinder\_binary)  
 SOHaloProfiles (haloFinder\_binary)  
 Output (dbi)

**SO Mass Function** x  
 Input Snapshot  
 o (csv)

**FOF Mass Function** x  
 Input Snapshot  
 o (csv)

**c-M relation** x  
 Input Snapshot  
 o (csv)

**c-M Emulator** x  
 Input Snapshot  
 output (csv)

**Power Spectrum** x  
 Input Snapshot  
 outfile (csv)

**P(k) Emulator** x  
 Input Snapshot  
 output (csv)