Harvesting E3SM Provenance

# Introduction

The first step to reproduce a simulation is to capture the necessary provenance information about the simulation. This quickstart guide explains the procedure in collecting provenance information from an E3SM simulation. The term harvesting is used to indicate the collection of already existing provenance information from simulations. The harvester script crawls through the simulation directory to collect the already identified important pieces of information about the simulation. The script creates intermediate csv files based on the collected information. There are three csv files generated and they are exported\_e3sm\_metadata\_msg.proven.csv, exported\_e3sm\_input\_files\_msg.proven.csv and exported\_e3sm\_output\_files\_msg.proven.csv. The next step is sending the csv files to ProvEn server.

# Prerequisites

* Python 2.7
* Access to the simulation directory

# Procedure

## Dependencies on other Quick Start Guides

## Steps

1. Run create\_provenance.py on the simulation directory

e.g.

python create\_provenance.py --e3smdir="/global/cscratch1/sd/golaz/ACME\_simulations/20171228.beta3rc13\_1850.ne30\_oECv3\_ICG.edison" --outputdir="/global/u2/b/bibiraju/rc13"

--e3smdir is the top level directory path of the E3SM simulation

--outputdir is the output directory where you want to save the generated provenance csv files.

## Next Quick Start Guide

Storing harvested provenance on ProvEn server Quick Start Guide

# Questions?

contact: [Bibi.Raju@pnnl.gov](mailto:Bibi.Raju@pnnl.gov), [Todd.Elsethagen@pnnl.gov](mailto:Todd.Elsethagen@pnnl.gov), [Eric.Stephan@pnnl.gov](mailto:Eric.Stephan@pnnl.gov)