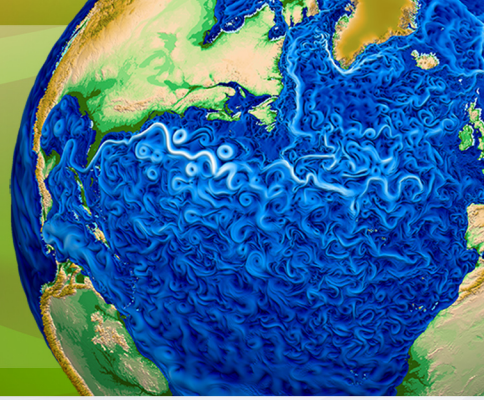
E3SM
Energy Exascale
Earth System Model



Energy Exascale Earth System Model (E3SM) Workflow

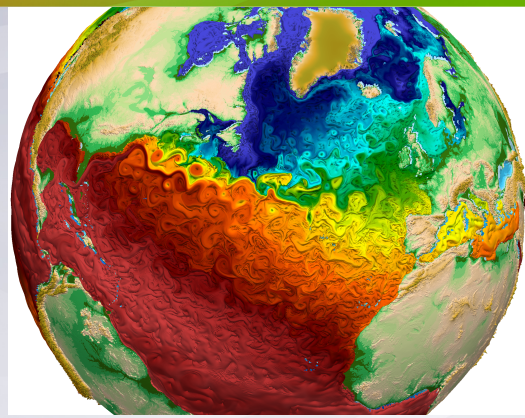
DOE Office of Biological and Environmental Research (BER) Project

Renata McCoy (E3SM Project Engineer),
Dean N. Williams, Valentine Anantharaj, (E3SM Group Leaders),
Dave Bader (E3SM PI and Council Chair)

ESGF F2F, Dec 5, 2017

LLNL-PRES-742657

U.S. DEPARTMENT OF ENERGY



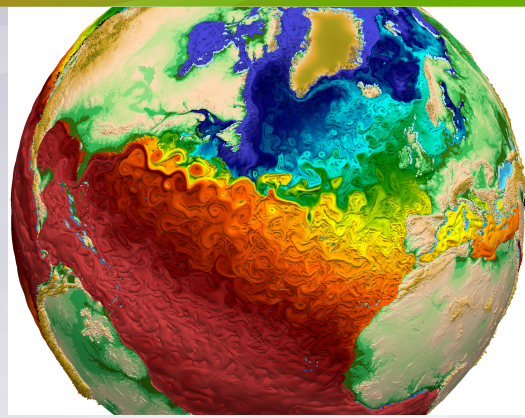
- Brief Introduction of the E3SM Project
- E3SM Development Cycle
- Timeline and Estimated Sizes of Data Production
- Key Difficulties Impeding Rapid Progress and Data Sharing
- Development Effort on End-to-End Workflow
- Automated Process Flow and Workbench
  - Data Management
  - Visualization
  - Diagnostic Packages
  - Provenance Capture
- Workflow Vision and Future Scenario

E3SM Energy Exascale Earth System Model

The E3SM Project is an ongoing, state-of-the-science Earth system modeling, simulation, and prediction project that optimizes the use of DOE laboratory resources to meet the science needs of the nation and the mission needs of DOE.

E3SM Project

Short Introduction



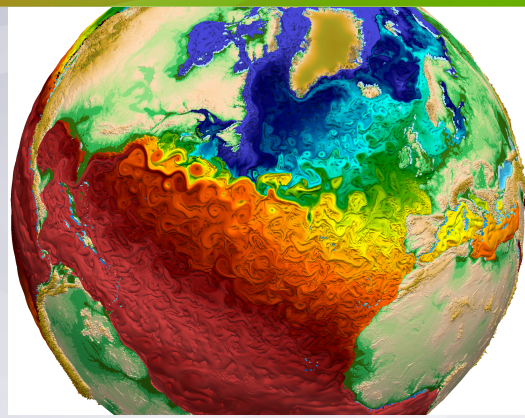
- Sponsored by DOE Office of Biological and Environmental Research (BER)
  - ~$20M per year
  - 8 Nat. Laboratories, University Partners
    - ANL, BNL, LANL, LBNL, LLNL, ORNL, PNNL, SNL
  - E3SM addresses critical, DOE mission-specific climate change questions
    - By developing Earth system and climate models at the leading edge of scientific knowledge and computational capabilities for exascale computing
    - By designing, executing and analyzing climate and Earth system simulations

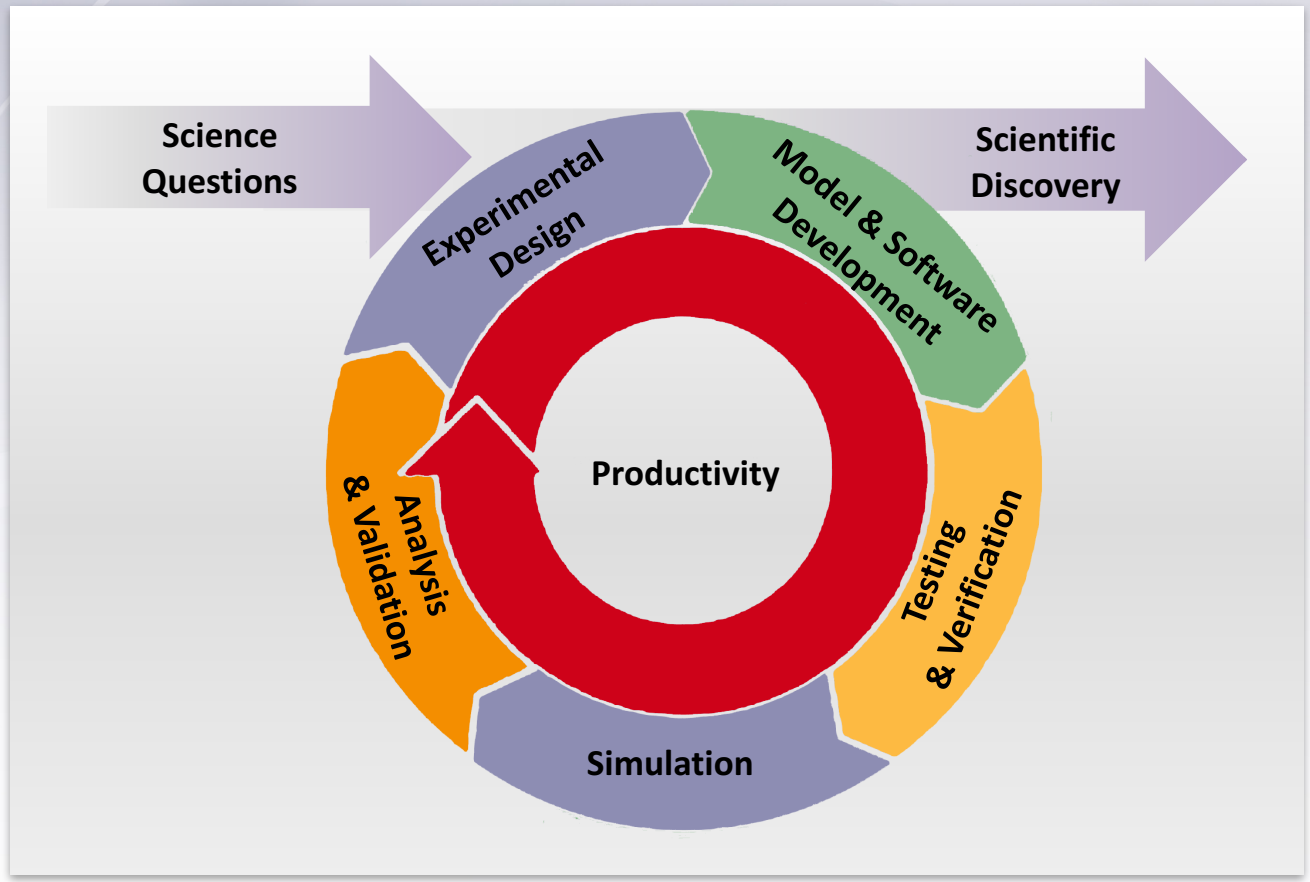
E3SM Energy Exascale Earth System Model

The E3SM Project is an ongoing, state-of-the-science Earth system modeling, simulation, and prediction project that optimizes the use of DOE laboratory resources to meet the science needs of the nation and the mission needs of DOE.

E3SM Development Cycle

End-to-End Workflow



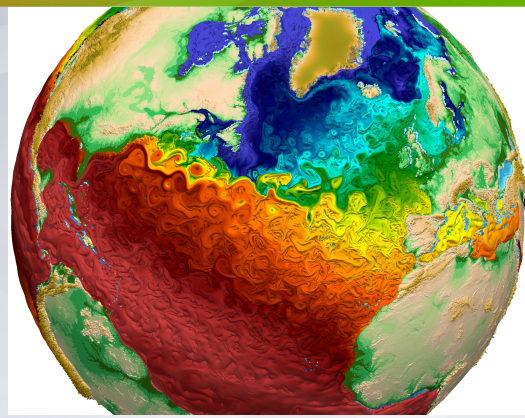


E3SM Energy Exascale Earth System Model

Workflow Group goal is to provide an automation of the end-to-end climate model development cycle

E3SM Target Simulations

Timeline and Estimated Sizes of Data



- E3SM version 1 Model & Data Release
  - target date **April 2018**

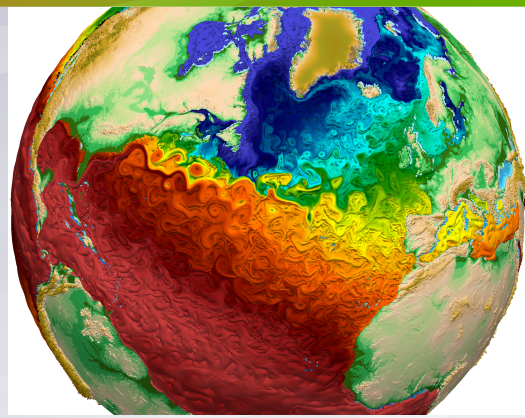
- Coupled Low-Res Runs (approx 1 deg)
  - CMIP6 DECK+
    - AMIP, Pre-industrial control, abrupt-4xCO2, 1%yrCO2, historical
    - 1400 to 1800 years --- **180 TB to 240 TB** (~13 TB/100 yr)
    - Monthly output **to be published to ESGF -- 70 TB to 90 TB**
    - CMORize after release date, then publish into CMIP6 ESGF Datasets

- Coupled Hi-Res Runs (approx ¼ deg)
  - Loosely follow HiResMIP
  - 50 years -- **210 TB.** (~4 TB/yr)
  - Monthly output to be published to ESGF – **60 TB**

The E3SM model is going to produce ~450 TB of data by April 2017 and will publish 150+ TB of data to ESGF.

Key Difficulties

Impeding Rapid Progress & Data Sharing

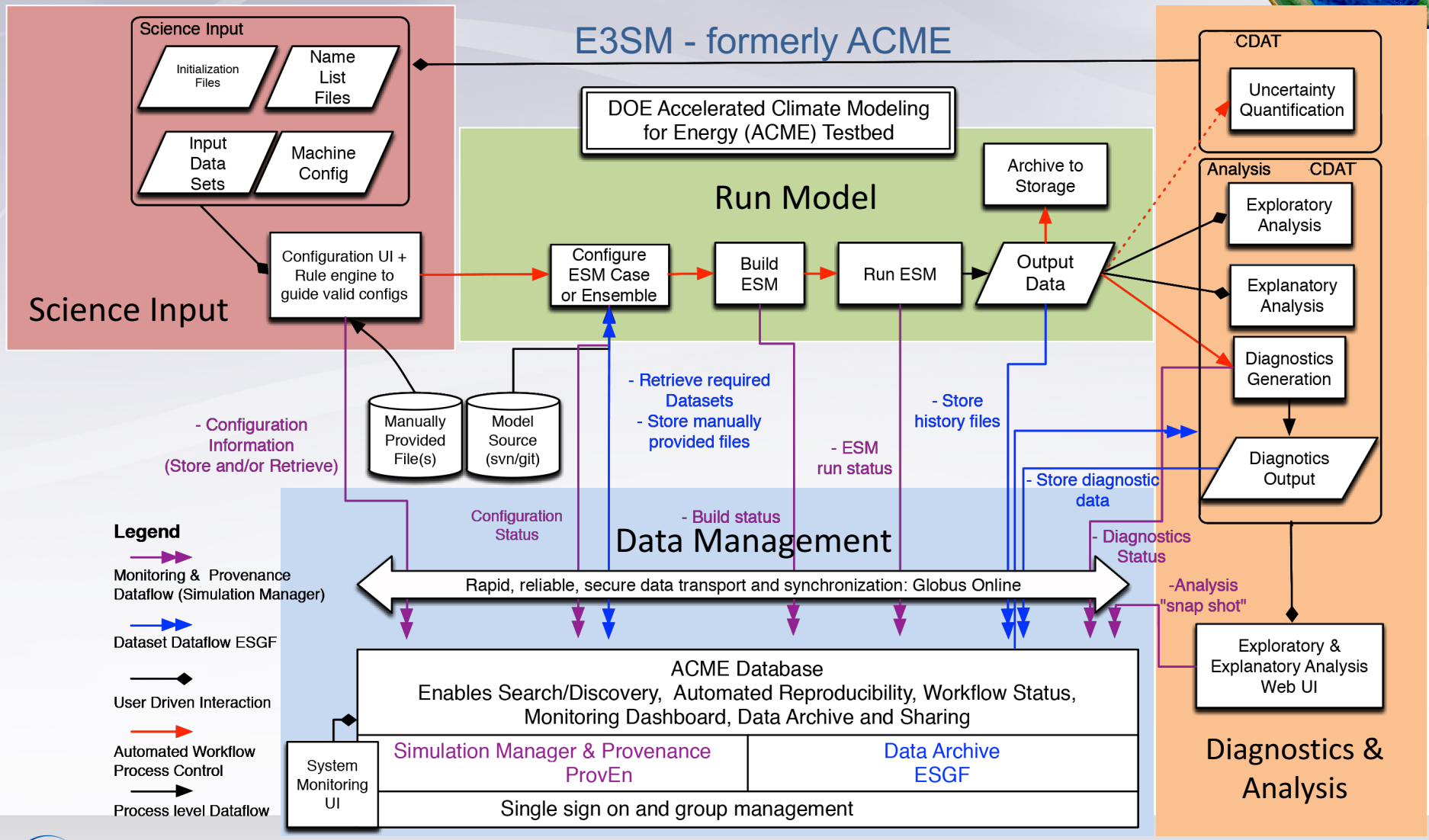


- Computational Time and Throughput
  - Computational awards through INCITE, ALCC, ERCAP
- Upcoming Disruptive Technologies
  - Unknowns about new “bleeding” edge machines
- Amount of data produced
  - Impacts further analysis, data sharing, moving, archiving
- Not enough automation in model run, analysis, evaluation, tuning, publishing, re-writing (CMORizing)
- Not enough integration between model development stages

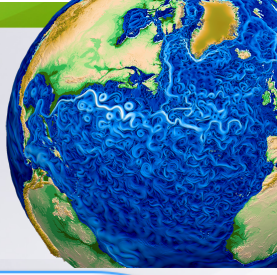
Workflow Group goal is to provide an automation of the end-to-end climate model development cycle

E3SM End-to-End Workflow

Detailed Diagram of End-to-End Workflow



E3SM automated process flow



HPC Edison

E3SM_run_script

CIME

ProvEn

Provenance Store

Short-term Archive

Zstash

Long-term Archive

Globus Transfers

ACME1 or AIMS4

Climatology and Timeseries File Generation

NCO

Diagnostics Generation

AMWG Diagnostics

A-Prime Diagnostics

Community Diagnostics Package

E3SM Diagnostics

PCMDI Metrics Package

ARM Diagnostics

Viewer

AMWG Viewer

E3SM Viewer

CDAT

Data Management

globus

Publish

Manage Data

Groups

Support

Submit: Select Collection

Lawrence Livermore National Laboratory

ESGF

ACME

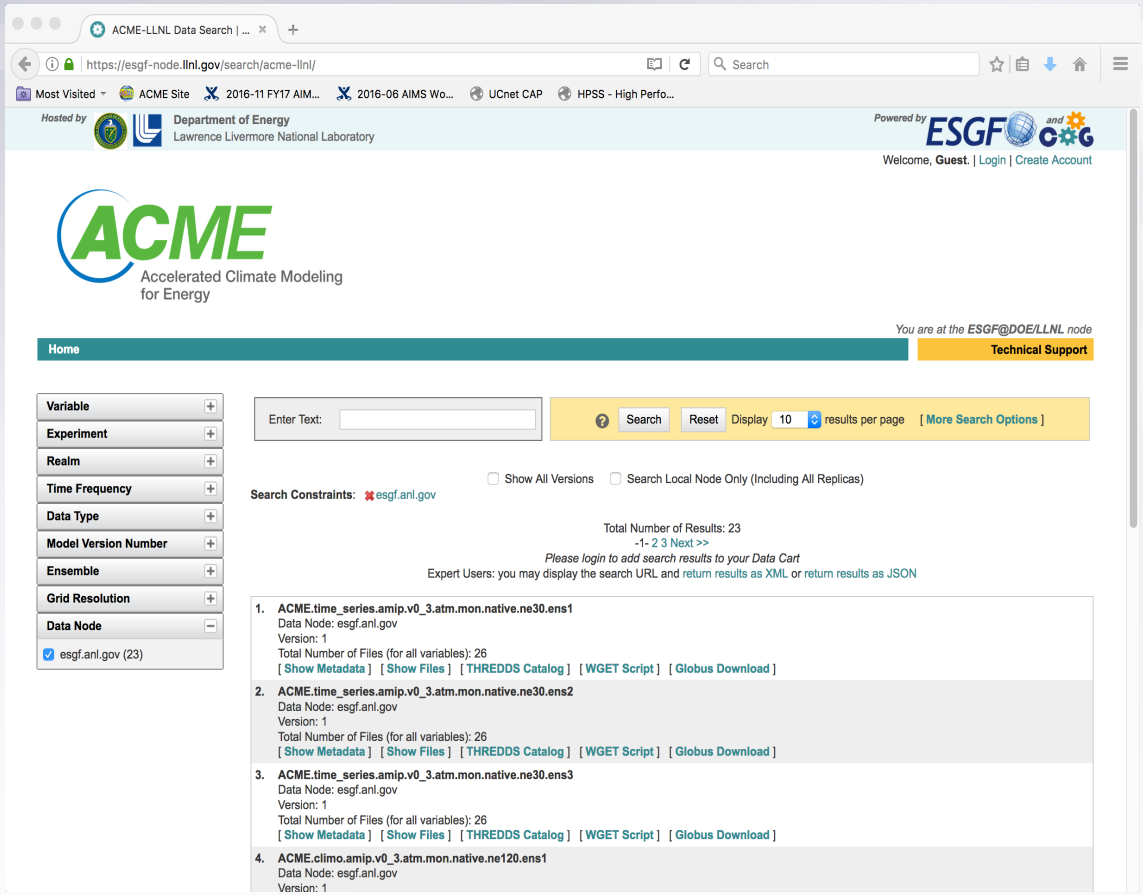
WCRP

CMIP5

E3SM Workbench

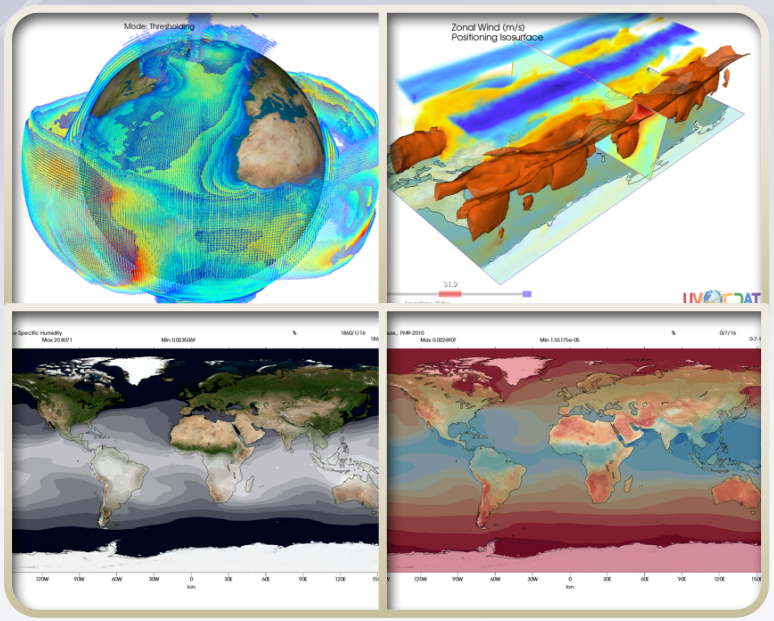
Data Management: ESGF, Globus

- Main publishing to **LLNL ESGF Node**
- Established of **ANL ESGF Data Node** as E3SM publication site and republicated 23 data sets formerly hosted at ORNL
- Support of publication process for CMIP6 metadata - needed for E3SM DECK participation in CMIP6
- Maintain **Globus** endpoints at E3SM sites at **ORNL, ANL, NERSC, LLNL**



Visualization: CDAT

E3SM's Analysis and Visualization: utilizes the Ultrascale Visualization Climate Data Analysis Tools (UV-CDAT)



Motivation

- The E3SM Workflow Group utilizes the Ultrascale Visualization Climate Data Analysis Tools (UV-CDAT), a Python-based tool designed to run ACME related analysis and visualization tools and techniques while capturing independent workflows for enhancing reproducibility

Features

- One-stop shop for analysis and visualization
- Easy-to-install via Anaconda (i.e., conda) open-source distribution
- Easy coupling with other libraries and packages, such as the latest regridding by interfacing to the Earth System Modeling Framework (ESMF) library
- Addresses projected E3SM scientific needs for data analysis and visualization

Impact

Examples for frequently repeated analysis:

- Seasonal climatology / anomaly
- EOF analysis
- Running average
- Hovmöller & Taylor Diagram
- Multi-model ensemble
- etc.
- Documentation via online manuals, tutorials, and Jupyter notebooks

Diagnostic Packages

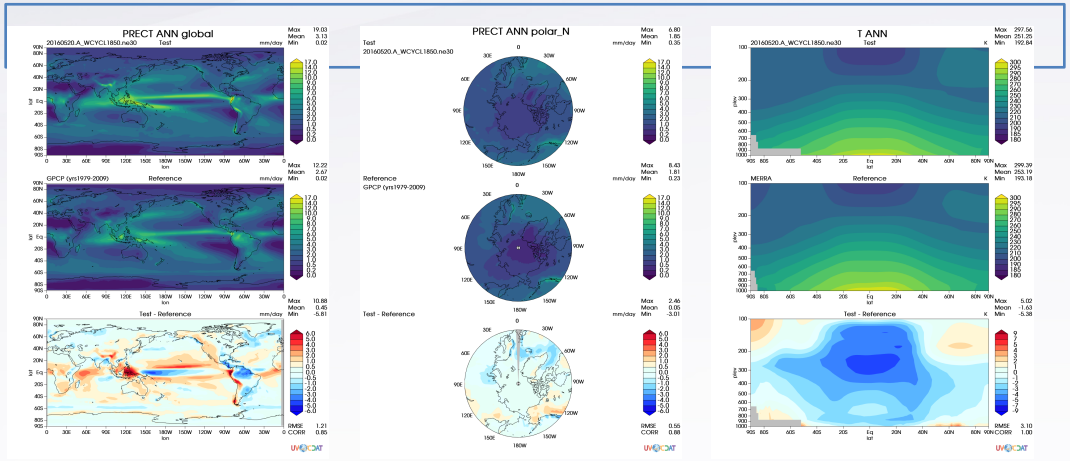
CDP Community Diagnostic Package: E3SM, ARM, PMP

Motivation

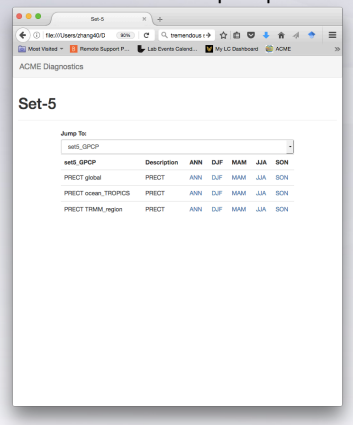
- A **modern, Python-based** diagnostics package for evaluating earth system models.
- Fully implement **datasets** and **diagnostics** developed by all of **E3SM science teams** and also NCAR's **AMWG** diagnostics sets.
- A **CDP-based** package to interact effectively with **PCMDI** and **ARM** diagnostics package.

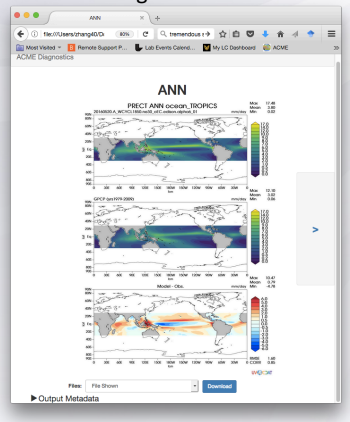
Diagnostics sets

- Lat-lon, polar, zonal mean and more.



Example of an E3SM Diagnostics run shown in the viewer: precipitation with different regions



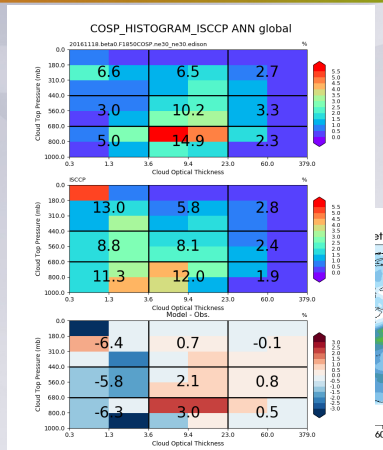


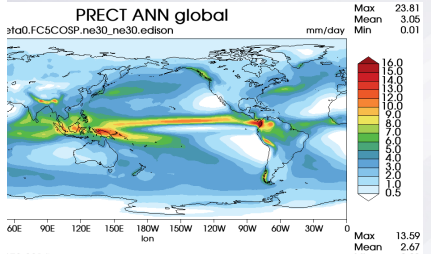
Features

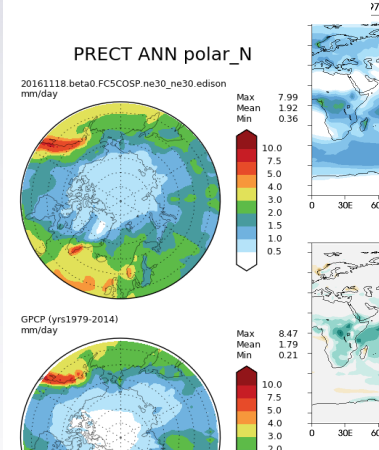
- Clean code base and simple design.
- Flexible configuration with many parameters to customize a run: select regions, pressure levels, multiple plotting back-ends, output format, color maps and more.
- Updated observational data repository.
- Documentation, demos, and user support.
- Software capability: continuous integration, test suit, and distributed computation.

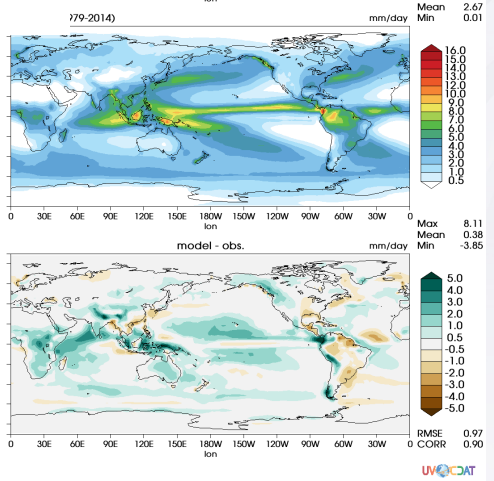
E3SM Diagnostic

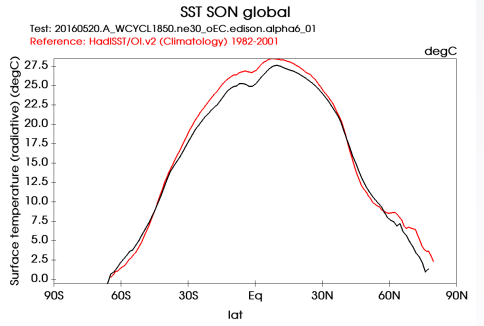
- Support for diagnostics based on seasonal or annual climatology data, including:
  - Latitude-Longitude contour maps (AMWG set 5)
  - Polar contour maps (AMWG set 7)
  - Zonal mean line plots (AMWG set 3)
  - Pressure-Latitude zonal mean contour plots (AMWG set 4)
  - CloudTopHeight-Tau joint histograms (AMWG set 13)
- Diagnostics for model vs obs., obs. vs obs. and model vs model, Jupyter note books as examples provided for each setting.
- Updated observational datasets available on LLNL and NERSC.
- Two graphical back ends: VCS and Matplotlib with cartopy.
- User–addable diagnostics during runtime.
- Enhanced color map configuration
- Serial and parallel run with multiprocessing or distributed.
- Documentation for LLNL and NERSC, detailed user's guide and examples.











ProvEn: Provenance Environment

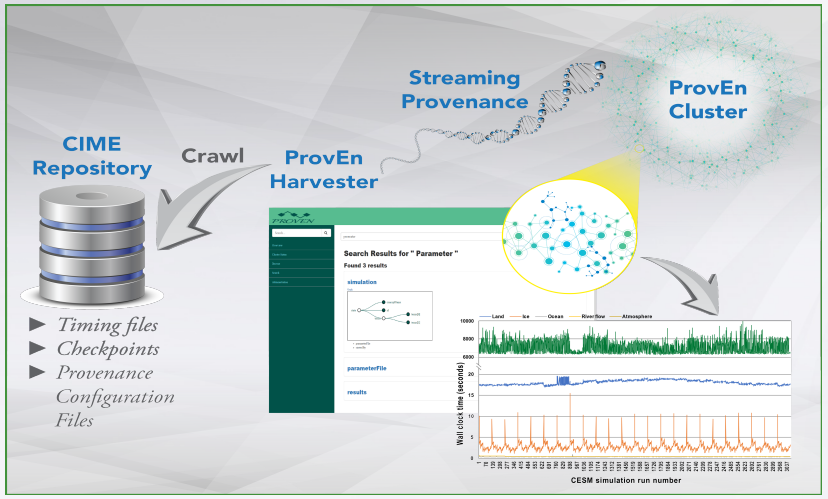
Motivation

- Investigate how workflow provenance can offer insights explaining what factors contributed to simulation results.
- Combine workflow provenance with performance metrics (for example timing data) for anomaly detection.
- Determine if reproducibility is possible by harvesting discrete information from previous E3SM simulation logs, configuration files, software descriptions, and compile settings.

Approach

- Harvest provenance information from E3SM case and run directories using ProvEn Harvester.
- The harvested provenance is stored in ProvEn's semantic store and visualized through ProvEn Dashboard and Jupyter notebooks.
- Integrate ProvEn with CIME to capture E3SM provenance is a unified way and use it for simulation reproducibility.

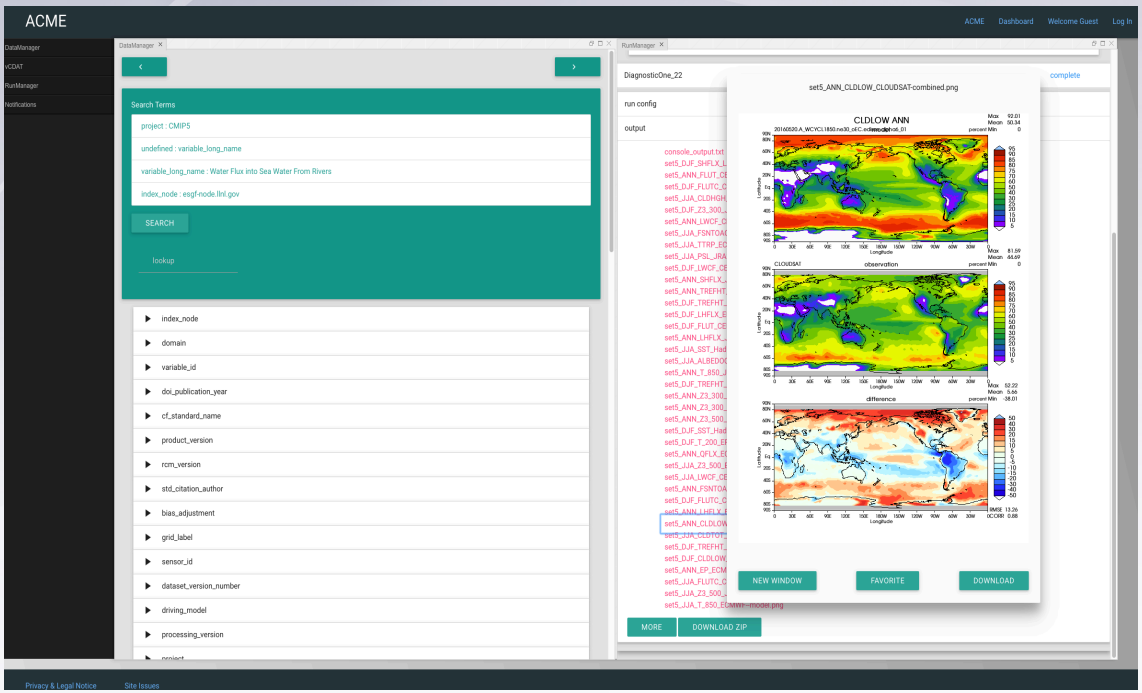
Provenance harvesting and visualization of E3SM simulation Data



Features or Impact

- Achieve simulation reproducibility by integrating with CIME
- Comparison of different E3SM runs
- Provide scientists running simulations performance and diagnostics tools.

E3SM Processflow and Workbench



Motivation

- Give the E3SM users access to a broad range of backend services, including regridded monthly climatologies and time series generation, multiple diagnostics (Atmospheric Modeling Working Group [AMWG], E3SM diagnostics, A-Prime diagnostics), automated Globus data transfers, and Earth System Grid Federation (ESGF) data publications
- The Workbench is a user interface that integrates the E3SM automated process flow tool to run a suite of post-processing steps automatically alongside the model.

Features

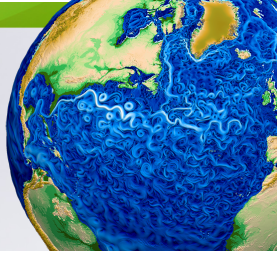
- The automated post-processing tool is used by E3SM scientists to run and generate all results automatically; giving users the ability to run a single command line script and be notified by email when the process has completed

Collaboration

- the Workflow Group had gathered a set of requirements from the E3SM Coupled Simulation Team and various model component team members

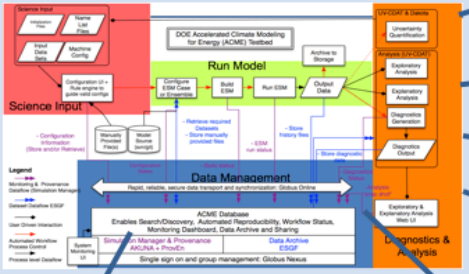
E3SM Workbench

User Interface Encapsulating Infrastructure



ACME Workbench – Overarching UI Encapsulating Infrastructure

Run End-to-End



CDP Community Diagnostics Package

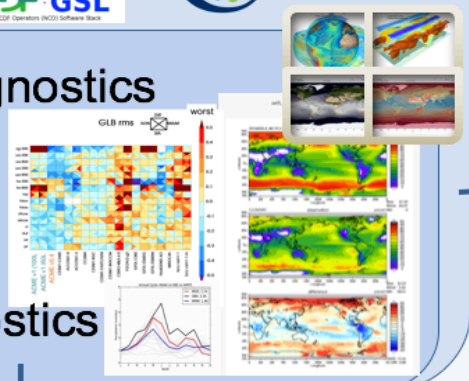
ACME Diagnostics

Accelerated Climate Modeling for Energy

PCMDI PMP

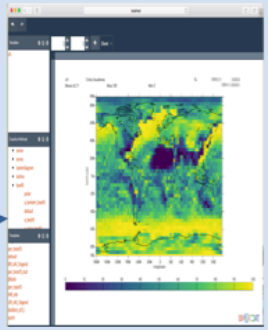
ARM Diagnostics

CLIMATE RESEARCH FACILITY



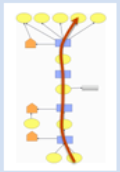
CDAT

VCDAT



Smart-client in Browser

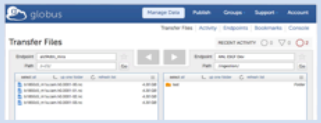
Provenance



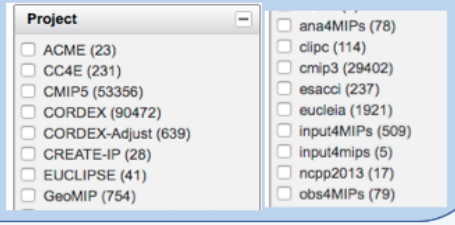
ProvEn Semantic Graphs

Data Management

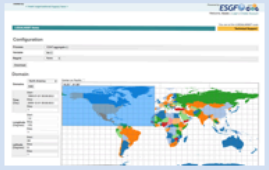
globus



ESGF Earth System Grid Federation



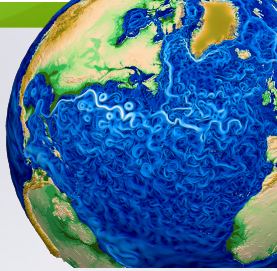
Server-side Computing



Data aggregation reduction

This integrated data ecosystem tightly interweaves every aspect of E3SM data research, from model development through interpretation and dissemination of research results.

E3SM Workflow Vision and Future Scenario



- The E3SM Workflow vision is to completely change the way people work on climate model development, diagnosis, evaluation and further analysis

- We want to trivialize the act of
  - Running a complicated model
  - Making simple changes
  - Reproducing and comparing the runs
  - Tracing provenance
  - Performing diagnostics and analysis
  - Creating additional user's derived analysis and variables
  - Archiving data and publishing results

- The end goals is for "non-experts" to be able to run the model, perform diagnostic, analyze the data and publish the results

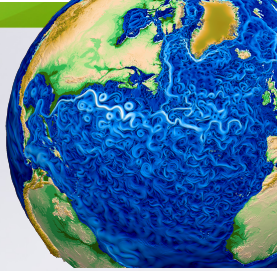
E3SM Workflow Group goal is to revolutionize the way people work with models and data

Future Scenario

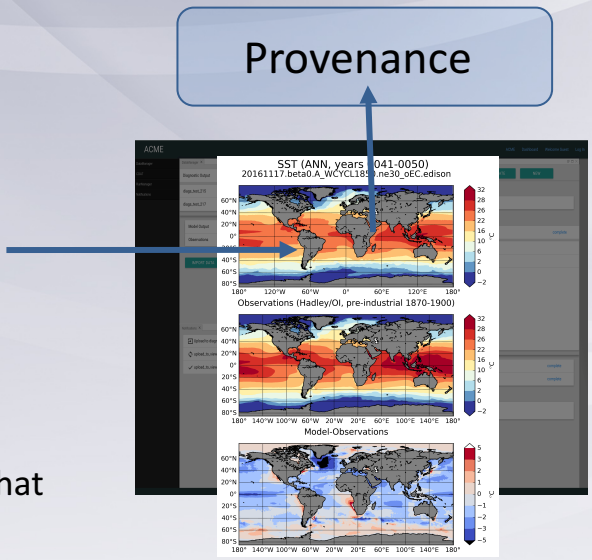
- Imagine you look at some diagnostic plot

E3SM Workflow Group goal is to revolutionize the way people work with models and data

Future Scenario

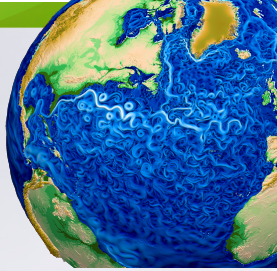


- Imagine you look at some diagnostic plot
- It knows its provenance
  - It knows the simulation that produced it,
  - It knows # tag of the GitHub code
  - It is connected to the data that produced the plot
  - It has the provenance (the code) of the diagnostic that produced it



E3SM Workflow Group goal is to revolutionize the way people work with models and data

Future Scenario

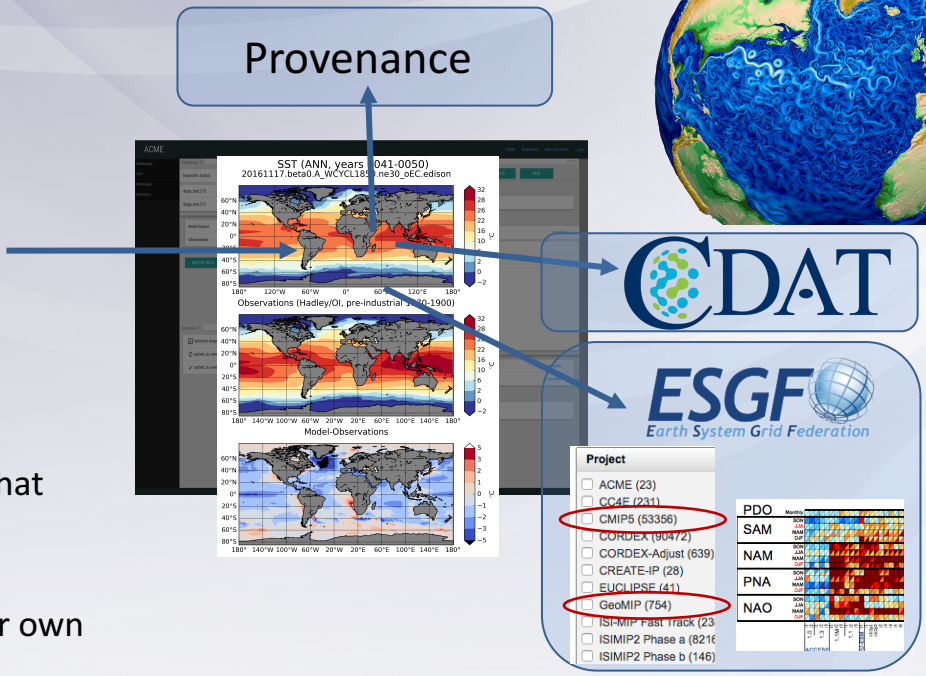


- Imagine you look at some diagnostic plot
- It knows its provenance
  - It knows the simulation that produced it,
  - It knows # tag of the GitHub code
  - It is connected to the data that produced the plot
  - It has the provenance (the code) of the diagnostic that produced it
- It is connected to CDAT (or vCDAT and other)
  - you can open the Python code and start coding your own analysis on top of the standard one or use UI

E3SM Workflow Group goal is to revolutionize the way people work with models and data

Future Scenario

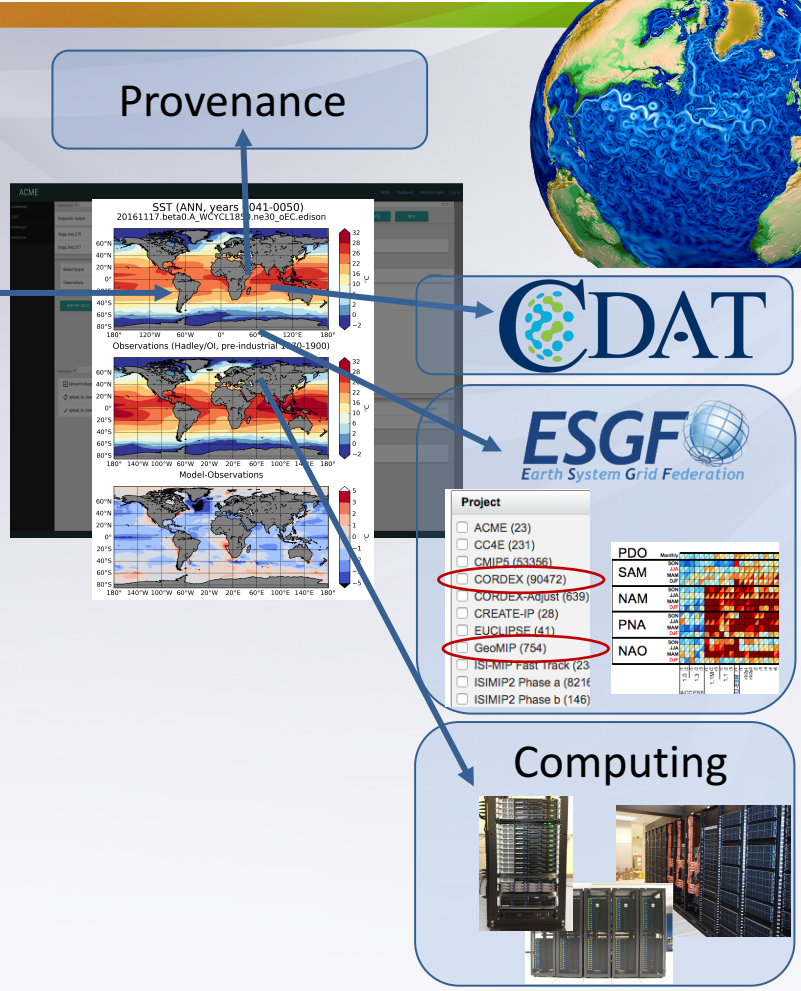
- Imagine you look at some diagnostic plot
- It knows its provenance
  - It knows the simulation that produced it,
  - It knows # tag of the GitHub code
  - It is connected to the data that produced the plot
  - It has the provenance (the code) of the diagnostic that produced it
- It is connected to CDAT (or vCDAT and other)
  - you can open the Python code and start coding your own analysis on top of the standard one or use UI
- It is connected to ESGF vast data archive



E3SM Workflow Group goal is to revolutionize the way people work with models and data

Future Scenario

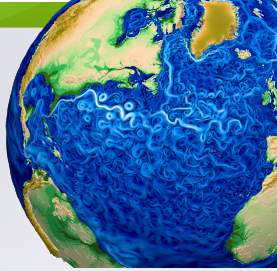
- Imagine you look at some diagnostic plot
- It knows its provenance
  - It knows the simulation that produced it,
  - It knows # tag of the GitHub code
  - It is connected to the data that produced the plot
  - It has the provenance (the code) of the diagnostic that produced it
- It is connected to CDAT (or vCDAT and other)
  - you can open the Python code and start coding your own analysis on top of the standard one or use UI
- It is connected to ESGF vast data archive
- It is connected to the smart, distributed computing resources
  - you can run your new analysis on the closest cluster



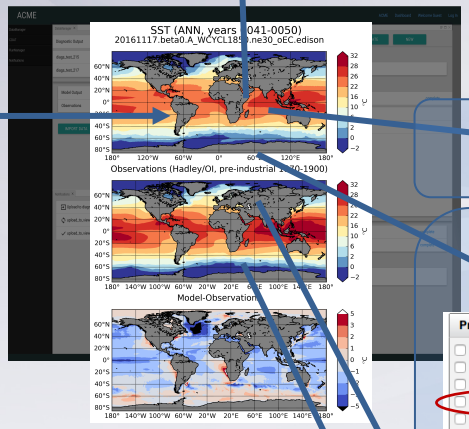
E3SM Workflow Group goal is to revolutionize the way people work with models and data

Future Scenario

- Imagine you look at some diagnostic plot
- It knows its provenance
  - It knows the simulation that produced it,
  - It knows # tag of the GitHub code
  - It is connected to the data that produced the plot
  - It has the provenance (the code) of the diagnostic that produced it
- It is connected to CDAT (or vCDAT and other)
  - you can open the Python code and start coding your own analysis on top of the standard one or use UI
- It is connected to ESGF vast data archive
- It is connected to the smart, distributed computing resources
  - you can run your new analysis on the closest cluster
- You can browse the code and make changes

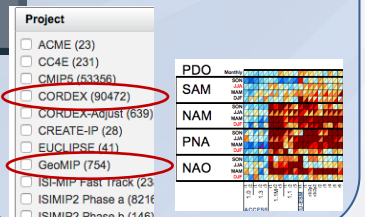


Provenance



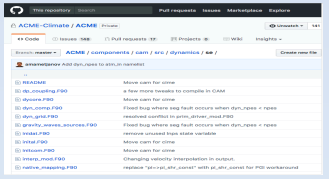
CDAT

ESGF
Earth System Grid Federation



Computing

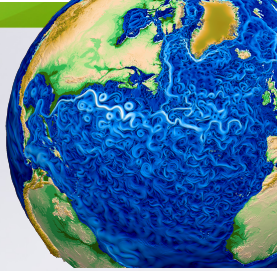
Code



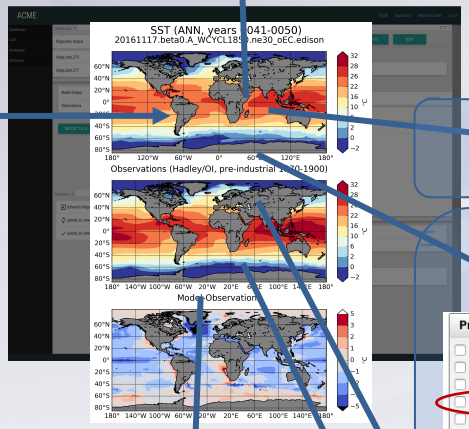
E3SM Workflow Group goal is to revolutionize the way people work with models and data

Future Scenario

- Imagine you look at some diagnostic plot
- It knows its provenance
  - It knows the simulation that produced it,
  - It knows # tag of the GitHub code
  - It is connected to the data that produced the plot
  - It has the provenance (the code) of the diagnostic that produced it
- It is connected to CDAT (or vCDAT and other)
  - you can open the Python code and start coding your own analysis on top of the standard one or use UI
- It is connected to ESGF vast data archive
- It is connected to the smart, distributed computing resources
  - you can run your new analysis on the closest cluster
- You can browse the code and make changes
- You can run the simulation with your changes
  - With some Livermore Computing (LC) resources in the back

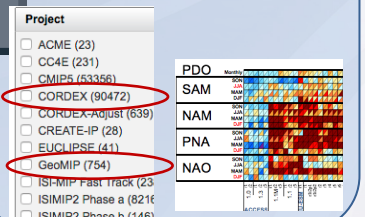


Provenance

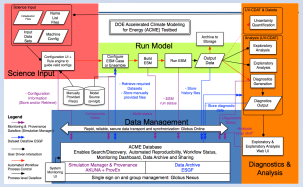


CDAT

ESGF
Earth System Grid Federation

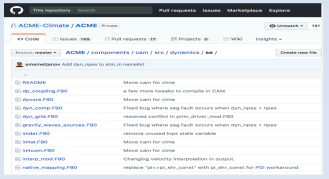


Run End-to-End



Computing

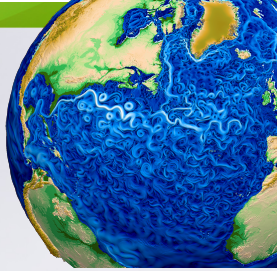
Code



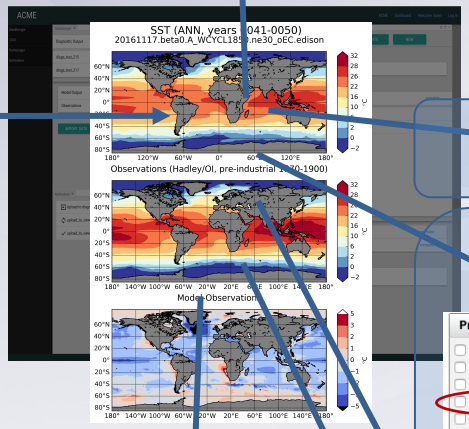
E3SM Workflow Group goal is to revolutionize the way people work with models and data

Future Scenario

- Imagine you look at some diagnostic plot
- It knows its provenance
  - It knows the simulation that produced it,
  - It knows # tag of the GitHub code
  - It is connected to the data that produced the plot
  - It has the provenance (the code) of the diagnostic that produced it
- It is connected to CDAT (or vCDAT and other)
  - you can open the Python code and start coding your own analysis on top of the standard one or use UI
- It is connected to ESGF vast data archive
- It is connected to the smart, distributed computing resources
  - you can run your new analysis on the closest cluster
- You can browse the code and make changes
- You can run the simulation with your changes
  - With some Livermore Computing (LC) resources in the back
- Come back in (for example) a week and check the resulting plot with your additional analysis

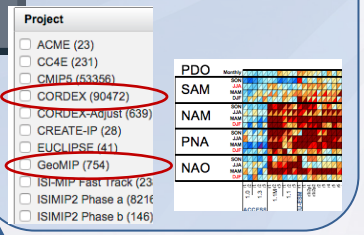


Provenance

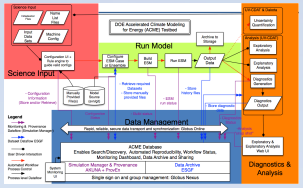


CDAT

ESGF
Earth System Grid Federation

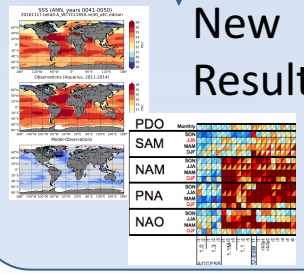


Run End-to-End

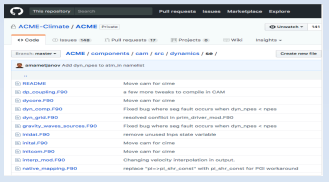


Computing

New Results



Code



E3SM workflow group goal is to revolutionize the way people work with models and data

Questions?