

# E3SM Data Management Plan

The E3SM project produces both software and simulation results. This plan describes the management and distribution of both types of data.

## Computer Software

E3SM software versions are considered developmental (development versions) until they produce final simulations that are analyzed and documented through a publicly released journal article or technical report (simulation document). Developmental versions are only available to project members and approved collaborators. Upon publication of the simulation document, source version of E3SM codes used to produce the simulations (open-source versions) will be available via an open-source license on GitHub (<https://github.com/>) as a "Public Project." We will use the tagging feature in the *git* version control system so that the precise code base is retrievable. The model version will be released together with the set of compsets, their input data, grids and maps needed to execute those released compsets. The software will be available for use under the license, but without support or consulting services beyond what is provided on the GitHub site. Open-source versions of the code will be maintained for five years from the date of release, or when the E3SM project ends, whichever is sooner.

## Simulation Output

All E3SM simulation output used in peer-reviewed journal articles and other openly distributed technical publications will be archived, maintained and curated for five years from the release of the publication, or when the E3SM project ends, whichever is sooner. As required by the Office of Science, data appearing directly in publications will be made publicly available through links on the E3SM public webpage (<http://climatemodeling.science.energy.gov/projects/accelerated-climate-modeling-energy>) associated with the publications. These data sets will be machine-readable in the self-describing netCDF format (<http://www.unidata.ucar.edu/software/netcdf/>), which is a broadly accepted good-practice standard utilized in the weather and climate science research communities.

The E3SM simulation output archive is expected to exceed several hundred tera bytes (TB). The E3SM project will maintain a catalog of all publicly released simulation data through the Earth System Grid Federation (ESGF) (<http://esgf.llnl.gov/>), with ESGF providing direct access to a subset of the data. Data catalog references to data that are not directly available will contain information on how to obtain the data from the project.

## E3SM Acknowledgement

**If you used the E3SM model or it's simulation data, please include the following Acknowledgment text in your resulting publication or presentation:**

"The E3SM model and simulation data were supported as part of the Energy Exascale Earth System Model (E3SM) project, funded by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research".

**Use the same statement for a publication, presentation, press release or similar document.**

**If referencing the E3SM in your publication, please add "formerly known as Accelerated Climate Modeling for Energy (ACME)" when first referring to the E3SM.**

**The E3SM data and model should be acknowledged in your paper and referenced with the appropriate DOI.**

The archival tags for the software should reference the DOI in Zenodo (<https://zenodo.org/>) (see for example: iESM [10.5281/zenodo.820079](https://zenodo.org/record/105281) or UVCDAT [10.5281/zenodo.592766](https://zenodo.org/record/592766))

**IMPORTANT:** The E3SM project has to write separate proposals for all of our computing resources. For these proposals (especially INCITE) it is important that we can cite papers which acknowledge the machines and computing facilities that were used. If a paper makes use of simulations that were run at ALCF and/or Anvil, NERSC, or OLCF, please be sure to include the appropriate acknowledgements:

ALCF: (Mira, Theta, Aurora)

The data were produced using resources of the Argonne Leadership Computing Facility at Argonne National Laboratory, which is supported by the Office of Science of the U.S. Department of Energy under contract DE-AC02-06CH11357.

Anvil:

The data was produced using a high-performance computing cluster provided by the BER Earth System Modeling program and operated by the Laboratory Computing Resource Center at Argonne National Laboratory.

NERSC: (Edison, Cori)

The data were produced using resources of the National Energy Research Scientific Computing Center, a DOE Office of Science User Facility supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

OLCF: (Titan, Summit)

The data were produced using resources of the Oak Ridge Leadership Computing Facility at the Oak Ridge National Laboratory, which is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC05-00OR22725.