Data Management Plan

For this project, we will create analysis packages that are applied to standard climate model output that is archived by the CMIP6 project as well as to observations, such as atmospheric reanalysis and satellite remote sensing datasets. The CMIP6 model outputs and observational data are already publicly available to the research community. As such, there is no need for us to make accessible those data. The methods of analyses that we will perform on these model output fields and observations will be explained explicitly in the papers that we will publish and scripts that we make available, so any derived quantities that we create from the original model output and observations will be reproducible by others. Key derived quantities from our analysis will be archived on the University of Washington's high-performance computing cluster Hyak beyond the length of the project.

The calculations and analysis methods that will be of broader use to the climate research community will be documented and made publicly available. Specifically, the proposed project will develop two primary products related to calculating the global flow of energy through the climate system: 1. Code for diagnosing energy transport from standard coupled climate model output and 2. Calculation of energy fluxes from high frequency re-analysis products and satellite observations. Both sets of calculations will be done separately for the zonal mean and zonal inhomogeneities and monthly and daily timescales. These calculations will be made available to the public via publicly accessible webpages at the University of Washington and will also be integrated into the Pangeo.io community platform for geoscience data as a gallery for community reuse. We do not expect any limitations on the distribution of our results due to privacy, confidentiality, security, or intellectual property issues. We have no requirements from other scientists choosing to re-use, or re-distribute our data.

Data Access and Sharing, Release Schedule: We expect an initial release of our observational calculations (with ERA5) and model code for zonal mean (including seasonal variations) in mid 2023. JRA and MERRA results will follow in 2024. The zonal inhomogeneous calculations are expected to be released in 2024-2025. Idealized model simulations will be released in 2023-2026 (in iterations).

Data Reuse: We expect a wide range of users for the data. Analysis of the global energy flow is ubiquitous in the climate research community spanning tropical precipitation, mid-latitude temperature and circulation and high latitude climate variability and long-term changes including sea ice melt and heat wave intensity.

Data Preservation: Model code and vital input/output data, and job scripts are backed up daily to secondary storage in the Department of Atmospheric Sciences at the University of Washington. Additionally, the data and code outlined in collections 1 and 2 below will be stored on the cloud and integrated into the Pangeo.io community platform for geoscience data as a gallery for community reuse.

Collection 1: Code for calculating atmospheric and oceanic energy transports from standard monthly mean climate model output.

Description: Code for computing global scale energy fluxes from standard model output. Separate scripts will be made for zonal mean and two dimensional (lat/lon) and annual mean, seasonal and interannual. data input to code will follow CMIP6 conventions and make use of the CMIP6 cloud computing

Format: *Source code: Python and* Matlab code.

Volume: 3.4 GB

Collection 2: Observational calculations of atmospheric energy transports and implied surface heat fluxes.
Description: Calculations of the vertically integrated atmospheric energy fluxes, the atmospheric column energy tendency and the implied surface heat fluxes constrained from satellite data. Zonal mean and 2-dimensional output at monthly resolution using NCEP, ERA65 and MERRA reanalysis (3 sets of calculations).
Format: NetCDF 4, data organization and convention will follow CMIP6 conventions (variable naming, time and space coordinates)
Volume: 40 GB per reanalysis set = 120 GB

Collection 3: Idealized climate model simulations
Description: Model simulations with altered evaporation coefficients and fixed atmospheric radiative heating profiles.
Format: NetCDF 4, data organization and convention will follow CMIP6 conventions (variable naming, time and space coordinates)
Volume: 20 GB per reanalysis set = 100 GB