

Intake / Pangeo Catalog: Making It Easier To Consume Earth's Climate and Weather Data

Anderson Banihirwe, Charles Blackmon-Luca, Ryan Abernathey, Joseph Hamman

Computer simulations of the Earth's climate and weather generate huge amounts of data. These data are often persisted on HPC systems or in the cloud across multiple data assets of a variety of formats (netCDF, zarr, etc...). Finding, investigating, loading these data assets into compute-ready data containers costs time and effort. The data user needs to know what data sets are available, the attributes describing each data set, before loading a specific data set and analyzing it.

In this notebook, we demonstrate the integration of data discovery tools such as [intake](#) and [intake-esm](#) (an intake plugin) with data stored in cloud optimized formats (zarr). We highlight (1) how these tools provide transparent access to local and remote catalogs and data, (2) the API for exploring arbitrary metadata associated with data, loading data sets into data array containers.

We also showcase the [Pangeo catalog](#), an open source project to enumerate and organize cloud optimized climate data stored across a variety of providers, and a place where several intake-esm collections are now publicly available. We use one of these public collections as an example to show how an end user would explore and interact with the data, and conclude with a short overview of the catalog's online presence.