

CESM Unified Postprocessing and Diagnostics (CUPiD)

Teagan King and Mike Levy
Mar 4th, 2024 - SEWG Meeting



GENERAL THOUGHTS FOR FLOW OF PRESENTATION - DELETE SLIDE LATER

- Mike [5 min]
 - ESDS Background
 - Vision of project / anticipated milestones
- Teagan [10 min]
 - Where we're at
 - A few things more relevant to SEWG:
 - i. Specific framework-related things we're addressing
 - ii. GeoCAT collaboration thoughts
 - iii. Collaborative work time thoughts?
 - iv. Any ways we may be able to improve CUPiD or strengthen collaborations with help from folks in this meeting?

About ESDS (Earth System Data Science)

Collaboration and grassroots effort started in Fall 2020 among NCAR/UCAR labs, centers, programs, and offices (LCPOs) focused on improving Earth system data science

- Led by a committee with representation from CGD, CISL, EOL, MMM, RAL, and Unidata
- Has been integral in the “push to python” (helping users transition from NCL to open source python packages)
- Community-building and creating a culture of open science and development



CESM Diagnostics and ESDS



- CESM users were hitting the limits of what NCL-based analysis could handle (and NCL is no longer being developed)
 - Advantage to making sure everyone replaced NCL with the same product
 - ESDS helped community find a solution
 - Pangeo stack: numpy, xarray, matplotlib, dask, etc
- Still had several independent efforts for analyzing CESM output
 - ADF, NBscuid, individuals writing one-off notebooks
- Next step: build a common framework
 - **CESM Unified Postprocessing and Diagnostics (CUPiD)**



ADF



CUPiD Collaborators

- **AMP:** Dani Coleman, Cecile Hannay, Brian Medeiros, Christina McCluskey, Jesse Nusbaumer, Justin Richling
- **CAS:** John Fasullo, Adam Phillips, Isla Simpson
- **CCR:** Gary Strand
- **CSEG:** Brian Dobbins
- **CESM:** Dave Lawrence
- **ESDS:** Katie Dagon, Teagan King, Mike Levy
- **ESMF:** Bill Sacks
- **GeoCAT (CISL):** Orhan Eroglu, Katelyn FitzGerald, Anissa Zacharias
- **OS:** Anna Deppenmeier, Gustavo Marques, Lev Romashkov
- **PPC:** Dave Bailey, Kate Thayer-Calder
- **TSS:** Sam Levis, Will Wieder

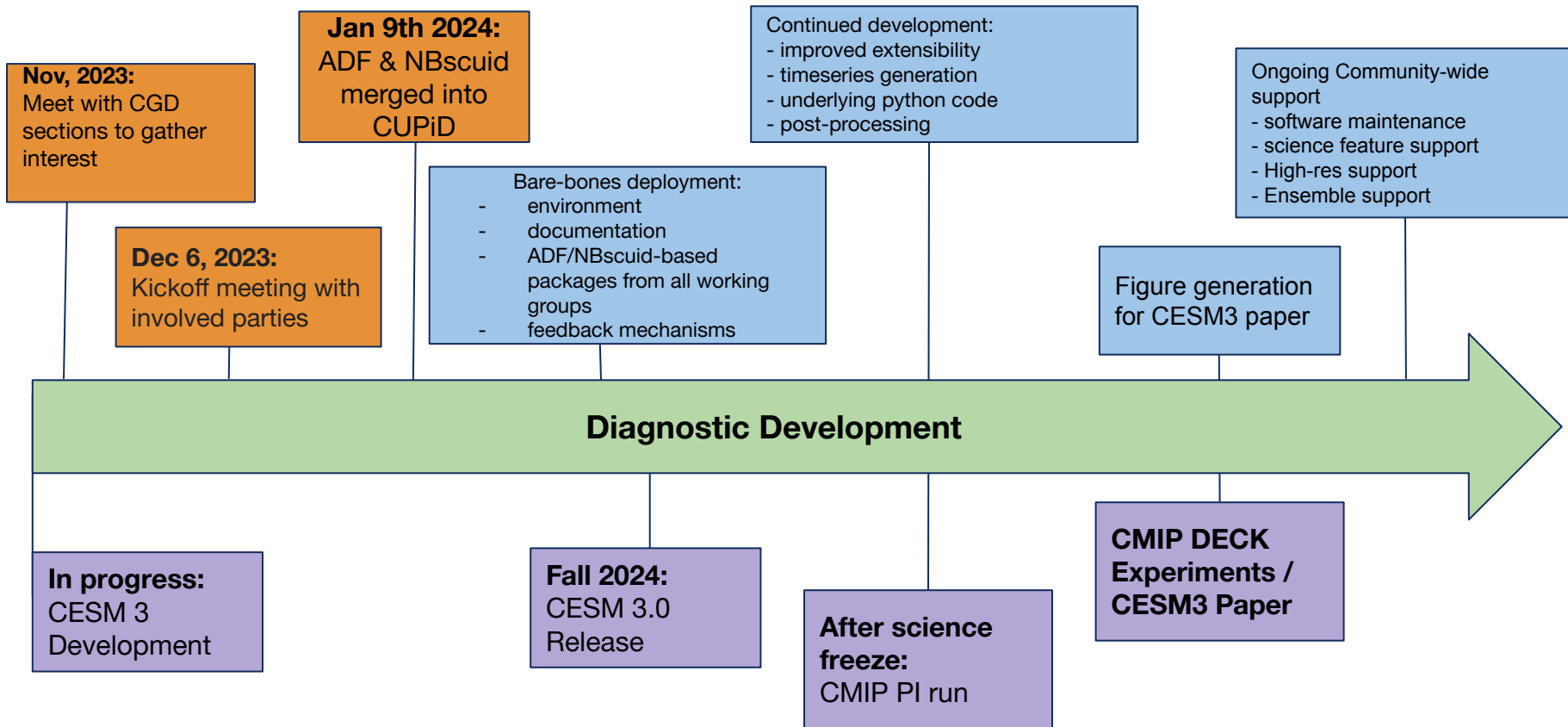
CUPiD Project Vision

A collaborative effort that unifies all CESM component diagnostics and provides

- Python code that
 - i. runs in an easy-to-generate conda environment, and
 - ii. can be launched via CIME workflow or independently
- Diagnostics for single/multiple runs and single/multiple components
- Ability to call post-processing tools that other groups are working on
- An API that makes it easy to include outside code
- Ongoing support and software maintenance

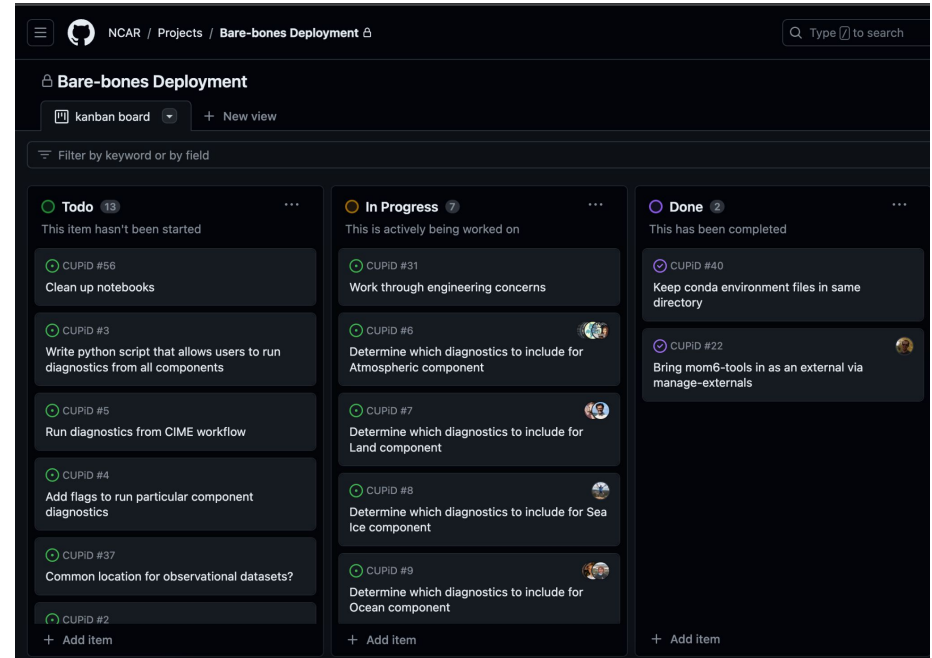


Anticipated Timeline



CUPiD Project Organization

- Open development on [github](#) (under NCAR organization)
 - [Documentation](#) on github.io
 - Fork / branch / PR strategy
- [Issues](#) and [Project Board](#)



Current State of CUPiD

- Single command line call to produce basic diagnostics for atmosphere, ocean, land, and sea-ice
 - User-editable **YAML file** controls what is run
 - Uses plumber (which wraps papermill) to run notebooks and pass parameters

```
(cupid-dev) $ cupid-run config.yml
index
adf_quick_run
ocean_surface
land_comparison
seaice
/glade/work/mlevy/conda-envs/cupid-dev/lib/python3.11/site-packages/plumber/dag/dag.py:455: UserWarning:
----- DAG render with warnings -----
----- NotebookRunner: index -> File('computed_notebook...ck-run/index.ipynb') -----
----- /glade/work/mlevy/codes/CUPiD/examples/nblibrary/index.ipynb -----
These parameters are not used in the task's source code: 'CESM_output_dir', 'lc_kwargs', and 'subset_kwargs'
----- NotebookRunner: adf_quick_run -> File('computed_notebook...df_quick_run.ipynb') -----
----- /glade/work/mlevy/codes/CUPiD/examples/nblibrary/adf_quick_run.ipynb -----
These parameters are not used in the task's source code: 'CESM_output_dir', 'lc_kwargs', and 'subset_kwargs'
----- NotebookRunner: ocean_surface -> File('computed_notebook...cean_surface.ipynb') -----
----- /glade/work/mlevy/codes/CUPiD/examples/nblibrary/ocean_surface.ipynb -----
These parameters are not used in the task's source code: 'subset_kwargs'
----- NotebookRunner: land_comparison -> File('computed_notebook...d_comparison.ipynb') -----
----- /glade/work/mlevy/codes/CUPiD/examples/nblibrary/land_comparison.ipynb -----
These parameters are not used in the task's source code: 'lc_kwargs', and 'subset_kwargs'
----- NotebookRunner: seaice -> File('computed_notebook...k-run/seaice.ipynb') -----
----- /glade/work/mlevy/codes/CUPiD/examples/nblibrary/seaice.ipynb -----
These parameters are not used in the task's source code: 'subset_kwargs'
----- Summary (5 tasks) -----
NotebookRunner: index -> File('computed_notebook...ck-run/index.ipynb')
NotebookRunner: adf_quick_run -> File('computed_notebook...df_quick_run.ipynb')
NotebookRunner: ocean_surface -> File('computed_notebook...cean_surface.ipynb')
NotebookRunner: land_comparison -> File('computed_notebook...d_comparison.ipynb')
NotebookRunner: seaice -> File('computed_notebook...k-run/seaice.ipynb')
----- DAG render with warnings -----
```

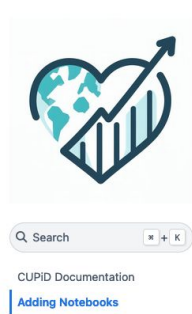
CUPiD run
command

CUPiD
configuration
file

Current State of CUPiD

- CUPiD provides two environments
 - cupid-run executable in cupid-dev (plan: have CISL add to NPL)
 - All notebooks run in cupid-analysis
 - Build on pangeo stack
 - Uses geoCAT packages
 - Active migration of functions to GeoCAT

- Step-by-step instructions for adding new diagnostics in this framework are [available online](#)



How to add diagnostics notebooks

Generally, a good fit for a diagnostic notebook is one that reads in CESM output, does some processing, and outputs plots, values, and/or new files (images, data, etc.) that are useful for evaluating the run.

1. Install the `environments/cupid-analysis.yml` environment (see [installation instructions](#)). Make sure that your notebook runs properly in this environment. If there are conflicts or missing dependencies, open an issue or talk to CUPiD developers so we can find a solution.
2. In your notebook, move all variables you might want to change (paths to data, dates to select, etc.) to a cell near the top. For example:

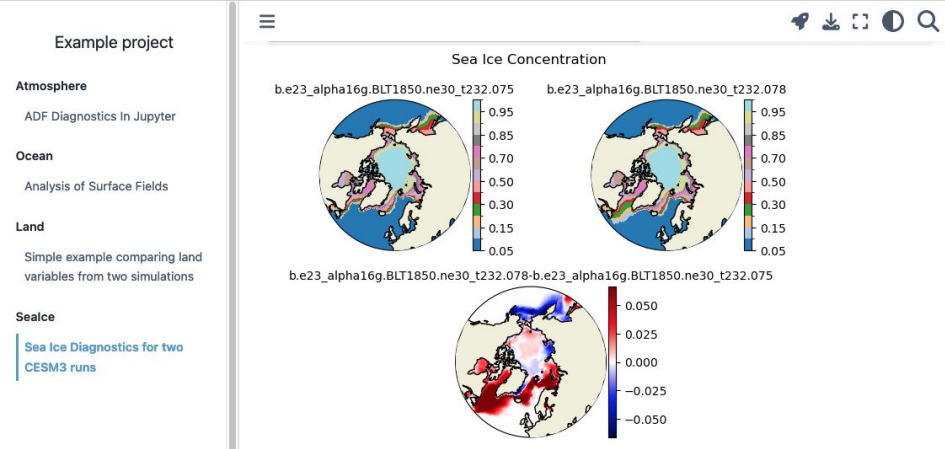
```
sname = "run_name"
data_path = "path/to/data"
dates = {"start_date" = "01/01/01",
        "end_date" = "01/01/02"}
```

3. Tag this cell as `parameters`. This means that when the notebook is executed by `CUPiD`, a new cell will be inserted just below this one with all of the parameters specified in `config.yml` (see step 5). To tag it, in Jupyter Lab, click on the cell and click the button with two gears in the top right ("Property Inspector"). Open "Common Tools." There, you can see a section called "Cell Tags." Click "Add Tag," and add one called `parameters` (exactly as written). **If you don't want to fully set up CUPiD, stop here and we can integrate the notebook into a CUPiD workflow from here.**

CUPiD Output

- CUPiD can use JupyterBook to turn notebooks into web pages
 - cupid-build function name likely to change

```
(cupid-dev) $ cupid-build config.yml
Running Jupyter-Book v0.15.1
Source Folder: /glade/work/mlevy/codes/CUPiD/examples/coupled_model/computed_notebooks/quick-run
Config Path: /glade/work/mlevy/codes/CUPiD/examples/coupled_model/computed_notebooks/quick-run/_config.yml
Output Path: /glade/work/mlevy/codes/CUPiD/examples/coupled_model/computed_notebooks/quick-run/_build/html
Running Sphinx v5.0.2
making output directory... done
```



Where we're headed

- Conda kernel available on Casper & Derecho
- Launch diagnostics from CIME workflow
- Port NCL-based code to python

Summary

- ESDS is a cross-laboratory initiative at NCAR
- CUPiD is an ESDS project focused on CESM3 (and beyond) diagnostics
- Still in early stages of development, but showing lots of promise