

# SEWG Discussion on Managing Model Output for CESM3

Discussion lead by

Gary Strand      Jim Edwards  
Sheri Mickelson      Jesse Nusbaumer

# First, let's talk logistics ....

- We'll be discussing two main topics
  - Data Standardization
  - Output Volumes
- We'll have about 80 minutes for discussion with a 10 minute break at about 3:40
- We recommend that you use the “raise hands” feature before contributing
- You can also contribute via the chat feature

# Data Standardization

- MIP compliance
- Standardizing model output across components
- Timestamps for averages
- Timeslice (“history”) vs. timeseries

# Data Standardization

- **MIP compliance**
- Standardizing model output across components
- Timestamps for averages
- Timeslice (“history”) vs. timeseries

Adding to and adopting controlled vocabulary used by the CF community (variable names, units, descriptions,...)

Using the same attributes and names between the components (where applicable)

MIP compliance is more FAIR (**F**indable, **A**ccessible, **I**nteroperable, **R**eusable)

Issues with conversion of ocean “cgs” base to “mks”

Translation to MIP compliance is error-prone

# Data Standardization

- MIP compliance
- **Standardizing model output across components**
- Timestamps for averages
- Timeslice (“history”) vs. timeseries

Abstract the output from the component models into a shared library.

Implementing asynchronous I/O?  
Hiding the performance cost may backfire

# Data Standardization

- MIP compliance
- Standardizing model output across components
- **Timestamps for averages**
- Timeslice (“history”) vs. timeseries

The “off by one” problem - Jan average reports 1 Feb time value

Recompute “time” based on average of “time\_bnds” for averaged fields

Instantaneous values denoted by absence of “cell\_methods” attribute only

# Data Standardization

- MIP compliance
- Standardizing model output across components
- Timestamps for averages
- **Timeslice (“history”) vs. timeseries**

Timeslice is faster, and thus better for model performance

Timeseries is more usable for the community and reduces bandwidth needs

# Output Volumes

- Controlling volume sizes
- Compression

# Output Volumes

- **Controlling volume sizes**

What's the minimum set to sufficiently characterize the climate of the simulation?

Default output is either zero or the minimum set; perhaps on a per-component basis

- **Compression**

All data for forever is impossible

LENS1 volume is ~550 TB; CMIP5 is ~180 TB; CMIP6 is ~550 TB

# Output Volumes

- Controlling volume sizes

Lossless compression (netCDF-4); saves about 55% total of space

Lossy compression is able to reduce storage by a factor of 5

- **Compression**

Lossy compression (Allison Baker, Haiying Xu), use CESM1-LENS as testbed?

# Funding and Support

How do we go about getting funding for these initiatives?

# Data Standardization

- MIP compliance
- Standardizing model output across components
- Timestamps for averages
- Timeslice (“history”) vs. timeseries

# Output Volumes

- Controlling volume sizes
- Compression