HDF5 use in Long Term Energy Modeling Systems at the U.S. Energy Information Administration (EIA)

HDF5 User Group Meeting 2023
Josh Whitlinger, Office of Integrated & International Energy Analysis
August 16, 2023 | Columbus, OH

Overview

- EIA Products

- Current use of HDF5 at EIA in WEPS

- Current work to get NEMS to use HDF5

- Future Work and Open Questions

What does EIA do?

The U.S. Energy Information Administration (EIA) is the statistical and analytical agency within the U.S. Department of Energy.

EIA is the nation's premier source of energy information.

By law, our data, analyses, and forecasts are independent of approval by any other officer or employee of the U.S. government.

EIA's analysis brings context and meaning to energy data

Short-Term Energy Outlook (STEO)

- Monthly forecasts of U.S. and global supply, consumption, trade, stocks, prices, and energy-related carbon dioxide emissions with a horizon of 12 to 24 months.

Annual Energy Outlook (AEO)

- Projects U.S. energy, supply, consumption, and trade over a 30-year period; the AEO is produced using EIA's National Energy Modeling System (NEMS)
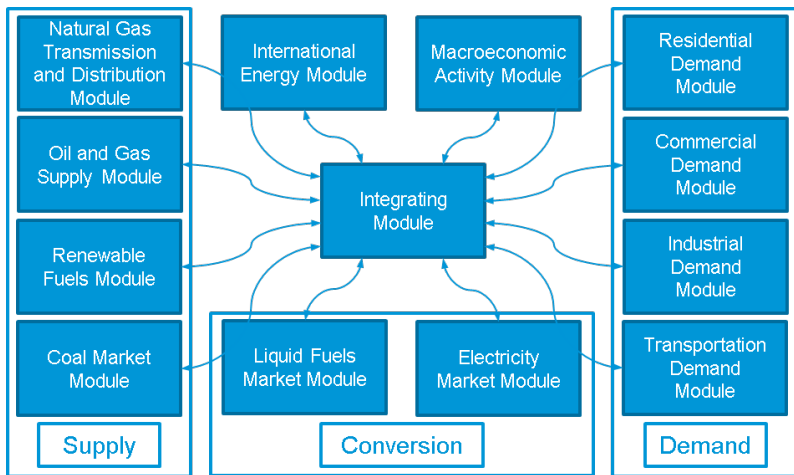
International Energy Outlook (IEO)

- Projects global energy supply, consumption, and trade over a 30-year period; the IEO is produced using EIA's World Energy Projection System (WEPS).

Other Reporting and Analysis

- *This Week in Petroleum* and *Natural Gas Weekly Update* summarize weekly developments in respective markets.

- *Today in Energy* summarizes topical energy trends, published daily on the EIA homepage

- Reports for Congress and key stakeholders

Models for AEO and IEO have similar, modular structures

National Energy Modeling System (NEMS)
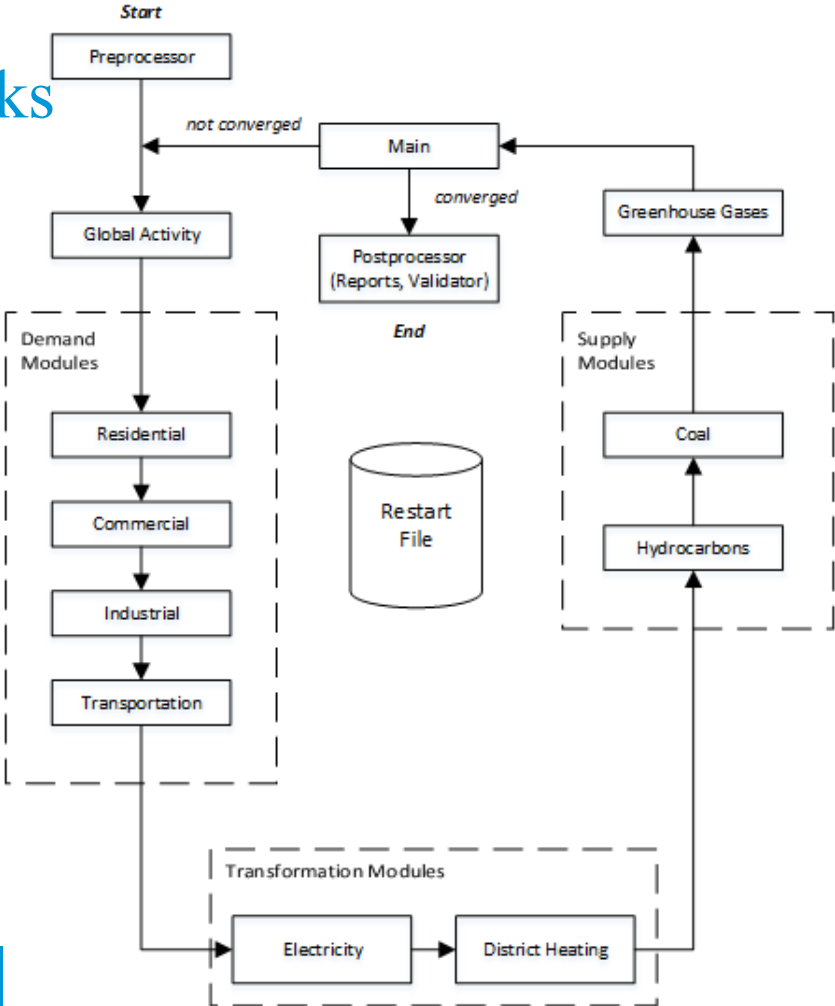U.S. model
used in Annual Energy Outlook

World Energy Projection System (WEPS)
global model
used in International Energy Outlook
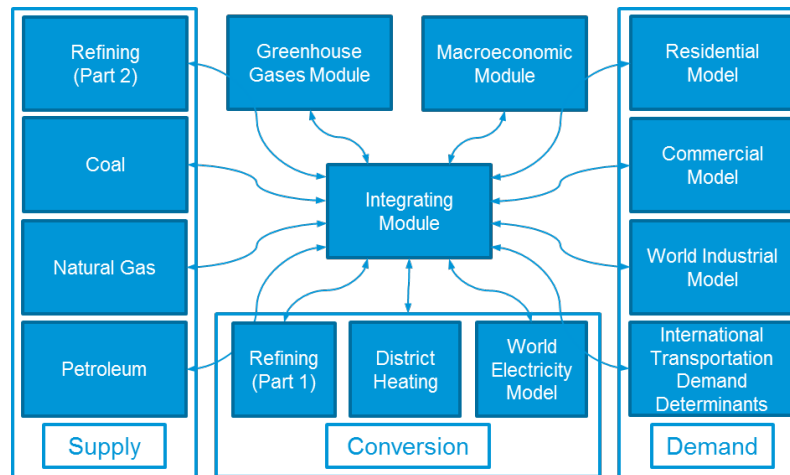
Current use of HDF5 in the World Energy Projection System
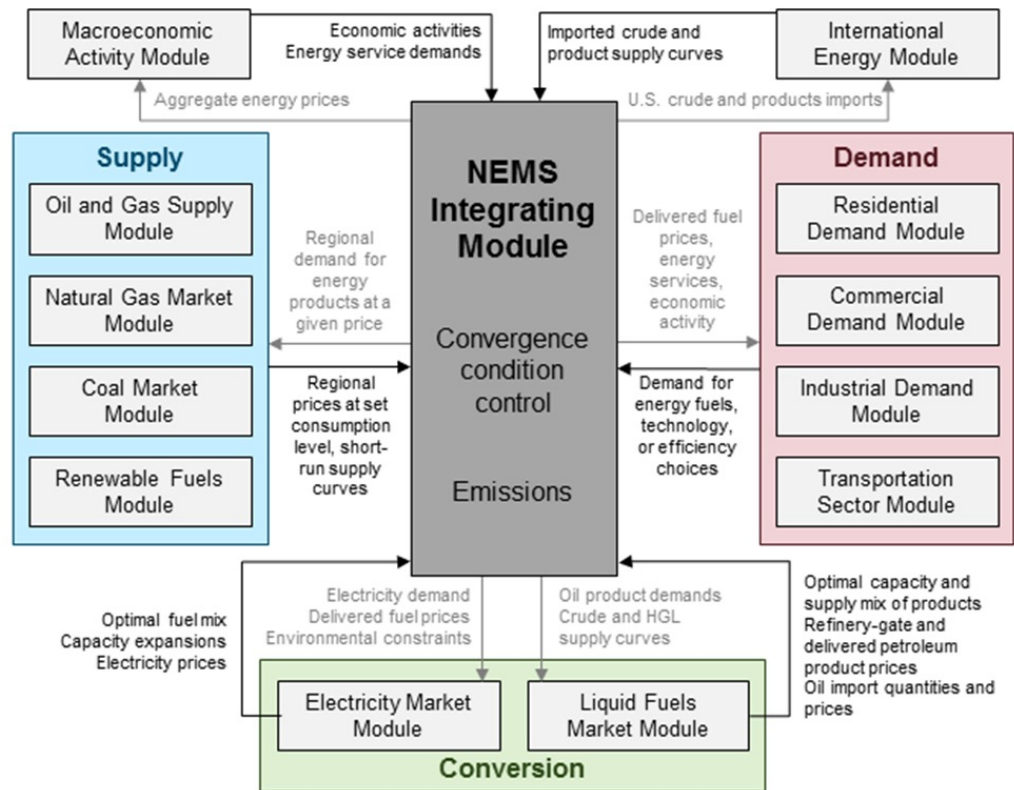# (WEPS)

How WEPS Works

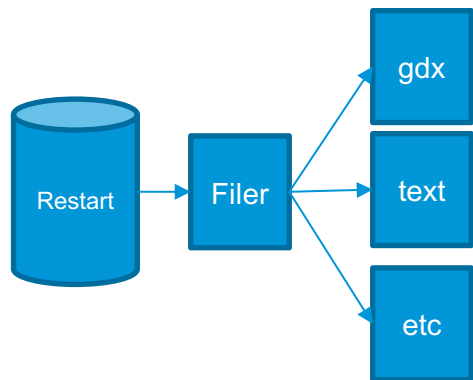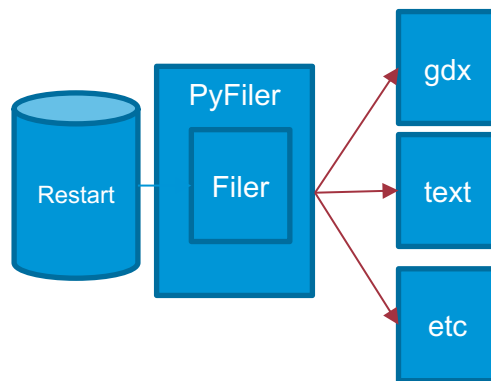# The National Energy Modeling System (NEMS)

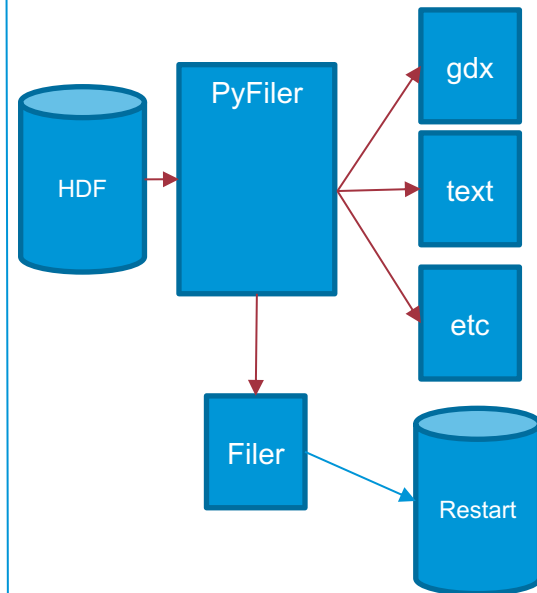# NEMS Structure

System diagrams

AEO2023

Now

Maybe the future?

Blue arrows are Fortran code | Red arrows are Python code

NEMS attempt with HDF

- Example for one variable:
    - In the EMM module, there exists some 5D variables that have the following dimensions:
        - 3 Regions
        - 12 Fuels
        - 61 Years
        - 4 Types
        - 2 Methods
    - Writing out this variable in HDF takes upwards of 20 seconds each time:
        - Read from Binary Restart File (1 sec)
        - Convert to multiple Pandas Dataframes (17 sec)
        - Send to HDF (1 sec)
        - we are required to write from unformatted binary restart file to HDF
    - Changing the order of the variable is a potential solution, but would require a lot of changes to the modules inside of Fortran
- So we're writing 4x2 pandas dataframes 2,196 times!

Future research questions

- Without editing the current existing i/o routines and structures in programs:
  - Are there methods to use to minimize I/O cost while still benefitting from the structured formatting of HDF?
    - Structuring of dataframes for HDF
    - Better way to do table structures with indices?
    - Minimize processing time
  - Exploring how changing the spine of NEMS to Python will impact these decisions