The Flexible Modeling System Runtime Environment (FRE)

October 19, 2007

Amy Langenhorst, RSIS/GFDL
Phone: 609 - 452 - 6572
Fax: 609 - 987 - 5063
Email: Amy.Langenhorst@noaa.gov

National Oceanic and Atmospheric Administration
Geophysical Fluid Dynamics Laboratory
Princeton, NJ 08542
http://www.gfdl.noaa.gov

Overview

- Why FRE? What is FRE?
- Mechanics of FRE
- Command line utilities
- Tour of the XML
- Database integration

Why FRE?

- Get the right source code files, versions
- Set the right parameters for compilation
- Set up the input, use correct versions of datasets
- Manage long production runs, job control
- Manage the output, postprocessing, figures
- Perform integrity testing on model configurations
- Make it simple to organize and share models
- Be able to reproduce your actions later

Key Features of FRE

- A standard mechanism for all FMS models
- A comprehensive description of model configurations
  - records details from code versions through compiling, running, and postprocessing
  - records model components, grids, dynamics and physics choices, all input parameters and data
- XML is self-documenting
- Easy to automate running regression tests

FRE Consists of...

- XML Model Description File: contains a complete description of one or more experiments
  - **cvs**: tags and modules needed to obtain code
  - **compile**: compiler flags, mkmf template, etc.
  - **input**: namelists, diag_table, datasets, etc.
  - **run**: simulation length, npes, etc.
  - **postProcess**: time series and averages, figures
- Command line utilities read XML file
  - **fremake**, **frerun**, **frepp**, etc

FRE Consists Of...

- CSH script templates
  - compile scripts
  - model runscripts
  - post-processing scripts
  - analysis figure producing scripts
- Standards and conventions for file names and locations
  - 'root' directory tree in /home
  - 'archive' directory tree in /archive

User Edits XML

```
<experimentSuite>
<setup>
  <directory type="root">$HOME/fms</directory>
  <directory type="archive">$ARCHIVE/fms</directory>
</setup>
<experiment name="cm2u">
  <component name="atmos">
    <cvs>
      <codeBase>fms_cm2</codeBase>
      <modelConfig>jakarta</modelConfig>
    </cvs>
    <input>
      <initCond file="/path/to/file"/>
      <gridSpec file="/path/to/file"/>
  …
```

User Calls frerun

```
> frerun -s cm2u
your job (432243) has been submitted
> frerun cm2u
TO SUBMIT: qsub /home/user/…/scripts/cm2u
>
```

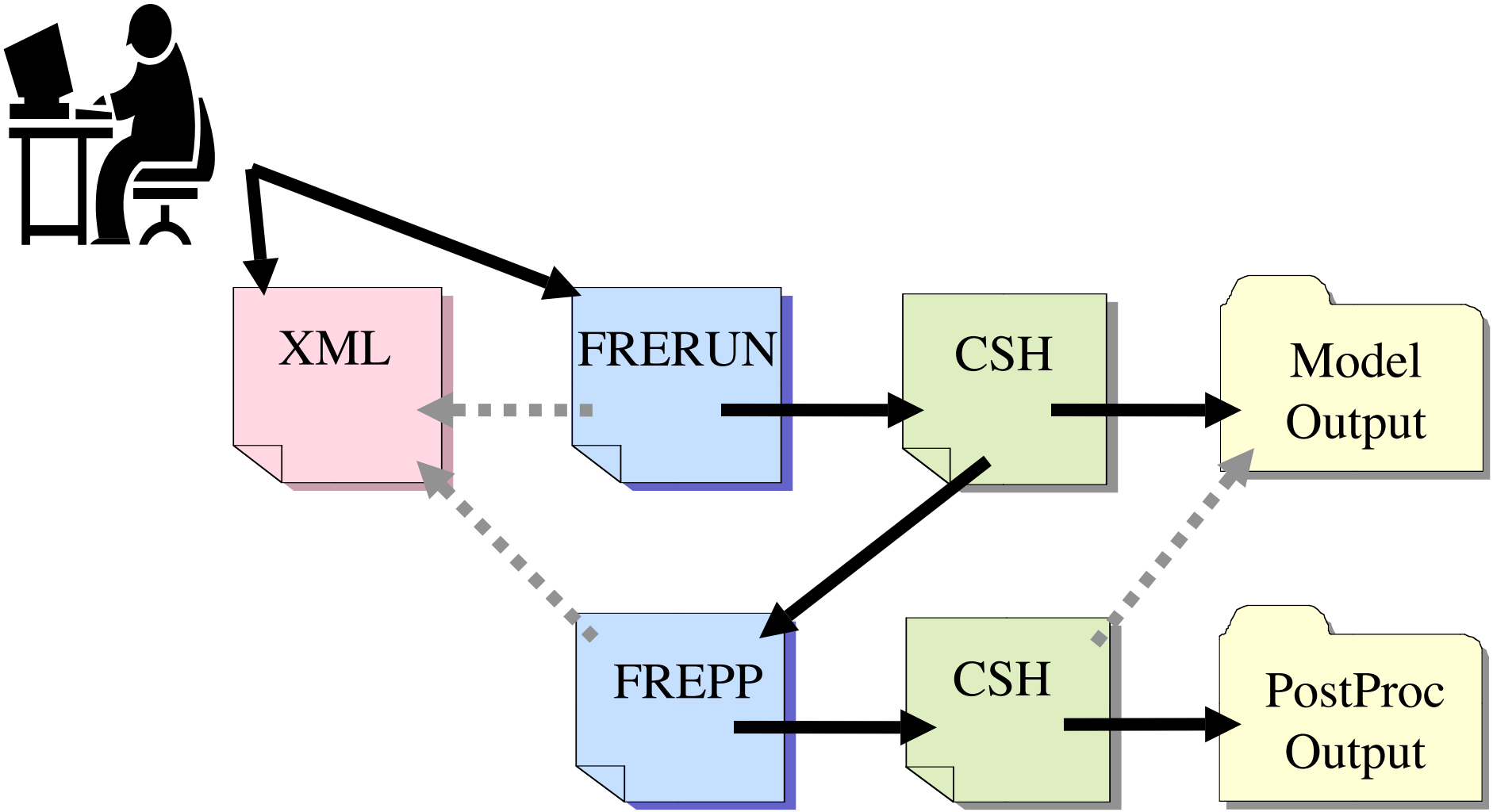
Model Output in Archive

```
cm2u
|-- ascii
|    |-- ascii_out.cpio
|    `-- stdout
|-- history
|    |-- 00010101.nc.cpio
|    |-- 00020101.nc.cpio
|    `-- 00030101.nc.cpio
`-- restart
     |-- 00010101.cpio
     |-- 00010201.cpio
     |-- 00010301.cpio
     |-- 00010401.cpio
     |-- 00010501.cpio
     |-- 00010601.cpio
     …
```

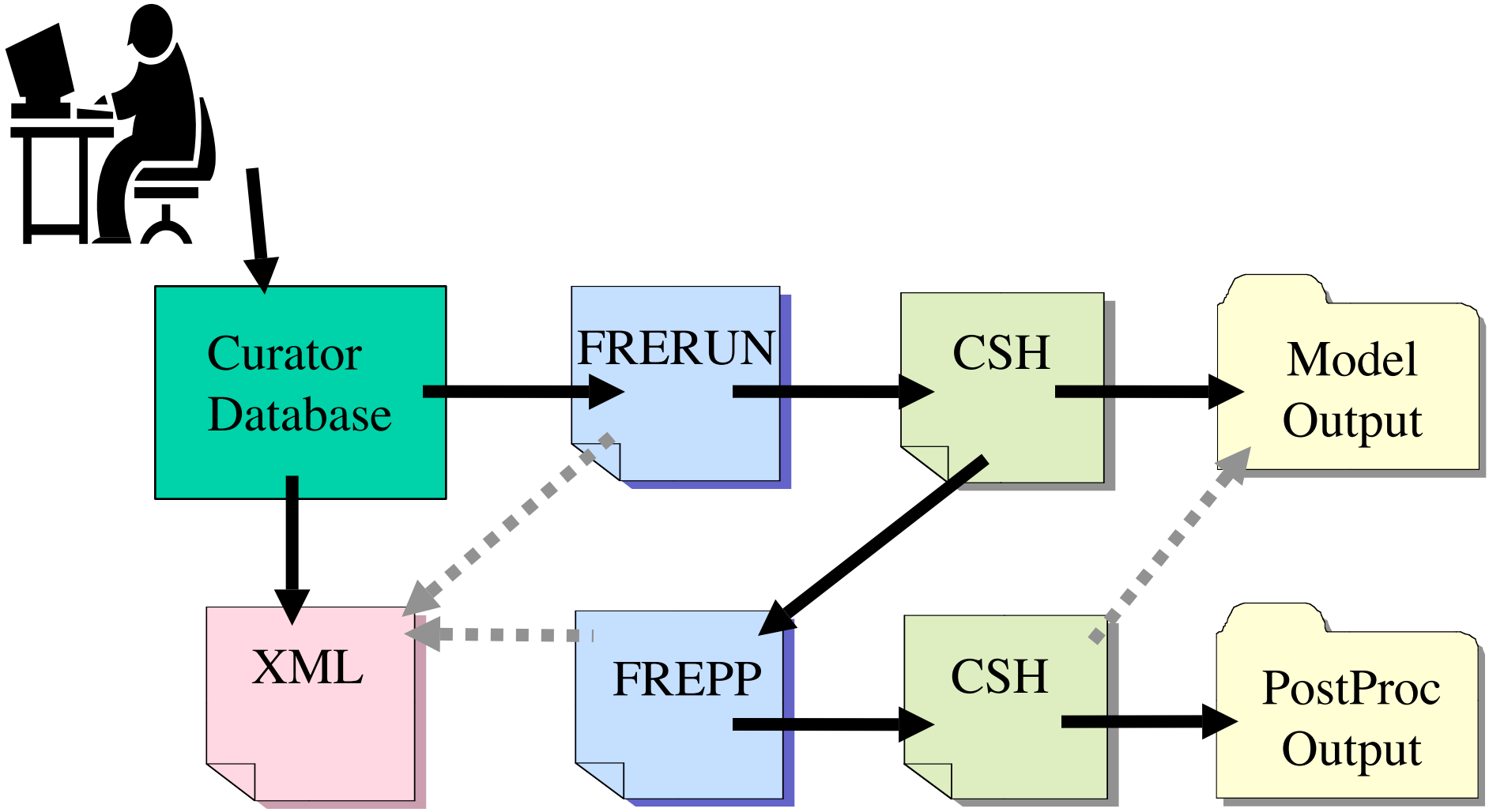
Postprocessing Output

```
cm2u
|-- ascii
|-- history
|-- restart
|-- analysis
`-- pp
    |-- atmos
    |    |-- ts
    |    |    `-- monthly
    |    |         `-- 5yr
    |    |              |-- atmos.000101-000512.mon.nc.cpio
    |    |              `-- atmos.000601-001012.mon.nc.cpio
    |    `-- av
    |         `-- monthly_5yr
    |              |-- atmos.0001-0005.01.nc
    |              |-- atmos.0001-0005.02.nc
    …              …
```

FRE Mechanics



FRE Mechanics



Auto-generated Figures

- A subset of the figures automatically generated for one of our publicly-released atmospheric models is available at http://www.gfdl.noaa.gov/fms/pubrel/analysis/m45_am2p14/
- Internally, users can browse experiments and their associated scripts, XML, data, and figures through our Model Development Database
- Many additional features coming with the new Curator Database

The FRE Utilities

- **fremake**: check out code, create/submit compile scripts
- **frerun**: create and submit run scripts
- **frepp**: create and submit postprocessing scripts
- **frestatus**: show status of batch compiles and runs
- **frecheck**: compare regression test runs
- **frepriority**: change batch queue information
- **fredb**: interact with Curator Database
- **frelist**: list info about experiments in an xml file
- **freppcheck**: report missing postprocessing files
- **frescrub**: delete duplicate postprocessing files
- **freversion**: use XSLT to change XML format

fremake

- **fremake -x xmlfile experimentname**
  - performs cvs checkout if you don't already have code in $rootdir/$experimentname/src
  - generates a simple shell script which calls mkmf (the makefile-generator) and make. You can qsub this script or run interactively.
- if you already have an executable, you do not need to call fremake

frerun

- **frerun -x xmlfile experimentname**
  - generates a cshell runscript which you can submit to the batch system or run interactively
- **frerun -r basic -x xmlfile experimentname**
  - **-r *keyword*** causes frerun to perform one or more short runs for reproducibility testing
- runscript will call frepp, the post-processing tool, as needed during the run

frepp

- called at the end of each model simulation year by the model runscript
  - creates and submits post-processing scripts to create time series, time averages, and analysis plots
- can be called ‘offline’ : outside of runscript
  - **frepp -t1983 -p4 -x xmlfile experimentname**
  - **-A** option: creates figures only, no data
  - **-d dir** specifies non-default history directory
  - **-c split** creates one job per component

frelist

- **frelist -x xmlfile**
  - lists all the experiments in an xml file and notes whether they inherit from another experiment
- **-v** option: also prints <description> tag contents for each experiment

frestatus

- **frestatus -x xmlfile experimentname**
  - prints the status of batch compiles and runs for the experiment
  - useful along with **qa -u user -n 45**
- **-c** option: status for batch compiles only
- **-r** option: status for batch runs only
- **-p** option: status for production runs only

frecheck

- **frecheck -x xmlfile experimentname**
  - verifies reproducibility by comparing the output of regression test runs with **resdiff**
  - prints timing statistics for the regression runs, generates a scaling plot
- **-l** option: just list all the output restart files
- **-L** option: list restart files with migration status (whether files are online or offline)
- **-n** option: perform no dmgets of offline files

freppcheck

- **freppcheck -x xmlfile experimentname**
  - checks for missing post-processing files
- will prompt you for the startYear and endYear that should be checked
- **-T** option: check all timeSeries variables
  - by default only the first variable is checked

frepriority

- **frepriority -x xmlfile experimentname**
  - manage job priority: queue allocations and job project
- **-a queue=num** change number of remaining queue allocations
- **-P proj** change job's project
- **-s** option: suspend job
- **-g** option: 'go' -- submit runscript

frescrub

- **frescrub -x xmlfile experimentname**
  - prints report of duplicated archive files for an experiment
- **-A** remove netcdf files when a cpio file exists
- **-B** remove redundant, uncombined 'raw' history files
- **-C** clean up $root by deleting excess old scripts
- **-D** all of the above cleanup options
- **-s** option: automatically qsub the script

freversion

- **freversion -v 3 in.xml out.xml**
- If there is a change in the XML language, users can use this program to automatically convert their XML files to the new version.
- Uses xsltproc

fredb

- **fredb -x xmlfile experimentname**
- to add or extract experiments easily to or from the experiment database
- in alpha, Sergey will demo

Comparing netCDF Files

- **ls -1 *.cpio | resdiff**
  - extracts each cpio file to a temporary directory and uses ***cmp*** to compare each restart file
  - if ***cmp*** fails, it uses a specialized utility ***nccmp*** to compare each netcdf file
- **nccmp file1 file2**
  - utility for comparing netCDF files
  - several options for omitting metadata comparison, excluding certain variables, etc

Work towards a new version

- Explicit support for model development ✔
- Support for generating multiple executables from a single source definition ✔
- Use of libraries to reduce compilation overhead ✔
- Greater flexibility in specifying model run units for both development and production (i.e. steps, seconds, minutes, hours, days, etc) ✔
- Script output tailored to intended use ✔
- Interpretation of dates so that datasets and other input can change with model date ~
- Support for ensembles
- Provide permanent record of production run information through integration with GFDL Data Curator ~
- Enhanced regression testing capabilities ~
- Enhanced post-processing capabilities; more control over time units of post-processing :(
- Support for FRE environment running on remote hosts ~
- Support for model implemented checkpoint / restart ✔

What else is new?

- Overhaul of Perl back-end
  - Aaron Watters
  - Saves user & "canonical" XML
    - "canonical" XML has all defaults filled in, is saved with a date stamp, is checked against for user changes
  - More object oriented, flexible and robust for implementing new FRE features

Issues we are facing

- How to organize the production of standard diagnostic figures
  - All scripts under CVS control?
  - XML entities with standard figure specs?
  - Wrapper scripts maintained by scientific groups?
- How can we give the users the shortcuts in the XML they desire without making things too complicated in other ways?
  - Inheritance, repeated information, XML variables

XML Tour

Where to go for more info

- **http://www.gfdl.noaa.gov/fms/fre**
  - FMS webpage -> **FRE Usage**
- An early beta version of FRE is available for public download at http://fms.gfdl.noaa.gov/
- **oar.gfdl.fre@noaa.gov** mailing list
- Amy.Langenhorst@noaa.gov