Developing the Earth System Modeling Framework as a standard

V. Balaji
Princeton University and GFDL

ESMF Advisory Board Meeting
Boulder CO
16 July 2004

Talk overview

- What are standards?
- Benefits of standards to our community
- Why ESMF is a good receptacle of standards
- The role of the standards body

What are Standards?

A *standard* is an agreed-upon common vocabulary and grammar to allow distinct communities that have identified common interests to communicate.

Examples of standards;

English "standardized" for the King James bible in the middle of the last millennium (Why?)

Microsoft proprietary, must be licensed to use.

Internet protocols open standard process, negotiated by the community, maintained by custodians.

Standards can be *descriptive* or *prescriptive*, and can carry implications of privileged communities.

But mostly, standards allow complex systems to be built in distributed fashion: components can be independently manufactured with a reasonable expectation of working reliably in conjunction with other such standard components.

The use of standards in software

The open source community provided a viable approach to the construction of software to meet diverse requirements for broad communities through "open standards". The standards evolve through consultation and prototyping across the user community. Software is designed as coupled independent components.

```
http://www.ietf.org/tao.html
http://www.ietf.org/rfc/rfc2026.txt
http://www.csrstds.com/openstds.html
```

Neither the Internet nor the World-Wide Web would have been possible without open community standards.

Benefits of standards to our community

In a complex and evolving technical and scientific environment, we propose the use of standards to transform the process of assembling, configuring, running and analyzing Earth system models from the current *heroic* mode to a *routine* mode.

Examples:

- If model data output followed standard name conventions for variables, it becomes easier to analyze data community-wide, and compare experiments from different models.
- If components of an Earth system model all use standard interfaces, it becomes possible to assemble diverse components into a model.
- If model grids and input parameters were described in a standard way, it becomes possible to share configurations between models.

Why ESMF is suitable for a community standard

- General standard-friendly features.

- ESMF classes and data structures are rich in descriptive metadata that can be seamlessly integrated into community data standards.

- High-level abstractions allow innovation.

General ESMF features

- ESMF is usable by models written in f90/C++.
- ESMF is usable by models requiring differentiability.
- ESMF is usable by models using shared, distributed or hybrid memory parallelism semantics.
- ESMF supports serial and concurrent coupling.
- ESMF supports multiple I/O formats (including GRIB/BUFR, netCDF, HDF, native binary).
- ESMF has uniform syntax across platforms.
- ESMF runs on many platforms spanning desktops (laptops, even!) to supercomputers.

ESMF's metadata-laden data structures

Earth system models can broadly be described as composed of components in which physical quantities are integrated on a physical grid. In a framework like ESMF, these are described in terms of 5 layers of abstractions consisting of *metadata-laden data structures*. These layers are:

`grid` describes the physical grid in a standard way, so that component-neutral regridding software can be used to transform quantities from one grid component to another, with no knowledge of those components themselves. We seek to inscribe the grid metadata within community standards and conventions, so that analysis tools cognizant of these conventions may take advantage of grid information.

`field` consists of the physical variable discretized on a `grid`, along with metadata describing the physical quantity itself. The field metadata in ESMF have been designed to resemble the CF convention, so that CF-compliant model output may be produced if desired.

`state` is the instantaneous state of some set of `field`s within a model component. Typically these are used as part of "import" and "export" states that are exchanged between components; but they are often used to contain the entire model state as well.

ESMF's metadata-laden data structures

`attribute` configuration attributes of a component: these are very generic, but are intended to contain all the physical input parameters used to configure a model.

`component` the top level entity of this design. Components are hierarchical: that is, they may be composed of other components. The top-level `component` is the application or model itself.

These software layers exist in the ESMF, and ESMF-compliant models in the near future will be using these abstractions, rich in metadata, to describe a wide range of models across the weather and climate community. Simply by using these abstractions and encoding them in model output, we are creating a layer of *formal, structured, hierarchical metadata*. We call this the *model metadata layer*, and it is the core of the proposed standard. The model metadata layer is what makes possible for either a fully-configured model configuration or a model dataset to be the result of a database query.

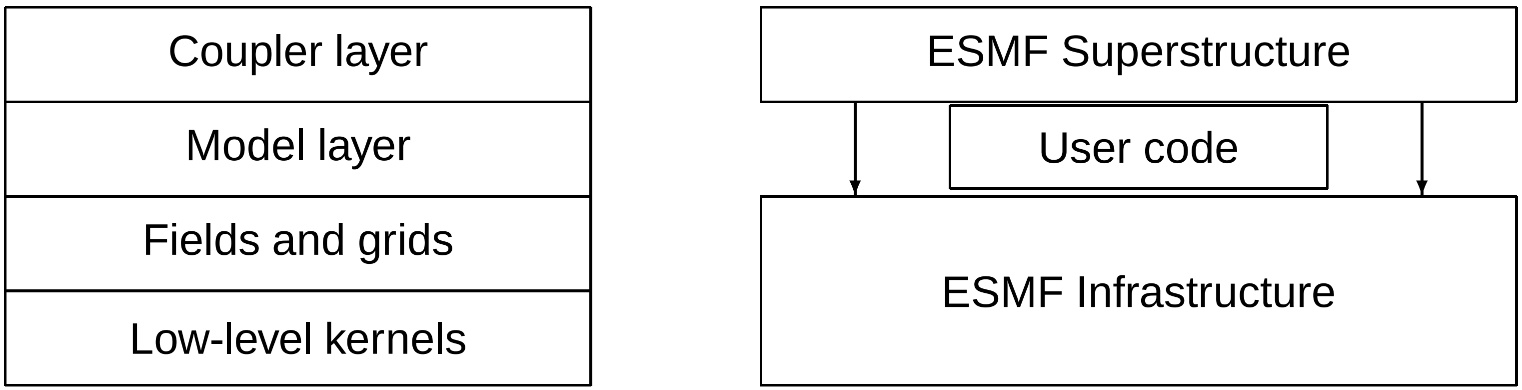
Beyond ESMF: linking model and data frameworks

Community data frameworks are under development. For model output data to be scientifically useful, the researcher must have some knowledge of how the data was produced. Model data requires a *model's eye view* description of the data, another layer of metadata, which includes:

- Description of model components: e.g FMS BGRID atmosphere, land and sea ice coupled to MITgcm ocean.
- Description of grid configurations and resolutions.
- Choice of physics packages and input parameters.
- Model state and its fields.

ESMF and PRISM are emerging standards that allow the development of the model metadata layer, based on the state data structures and its base classes. Modeling framework data structures map directly on to community hierarchical metadata. Observational data has an analogous data structure within ESMF as well: the *location streams* used in data assimilation.

Upward evolution of standards



Standards currently sit in the low-level *machine layer* (e.g MPI, netCDF).

By developing an *open standard* for the *grid layer*, we permit much greater freedom of innovation in software and hardware architectures for scalable systems.

The “standard benchmarks” (LINPACK, SPEC, etc) do not yet reflect this trend.

Potential use scenarios

Climate scientists setup (assemble components, configure input parameters); comparisons (run configurations, results, with data); branch runs, ...

Impacts studies query models by pattern, couple biogeochemistry model either offline with dataset or online with model.

IPCC, MIPs descriptions of intercomparisons, setup new MIPs, archive MIP results.

Policymakers, industry and educators High-level access to swathes of model data.

Publication link datasets to publications; introduce interactive aspect to publication; annotation of data, certification and quality control.

Portability automatic best-practice configuration appropriate for platform.

Operations higher rate of technology transfer from research to operations.

Guidelines for an ESMF standards body

- The community is the worldwide Earth System modeling community. Liaison activities with the Earth System data community, and broader scientific domains (space weather, geosciences...)

- The standards body will be constituted independent of any implementation of ESMF, including the core implementation.

- Membership in the standards body to be open and voluntary.

- The charge of the standards body is to take custody of the ESMF standard on behalf of the community; to maintain a set of interfaces that can be implemented on current architectures, and extensible to future architectures; energize and provide input to the development of Earth system data standards (CF, GO-ESSP).

- The body will work with benchmark groups (SPEC) to get standards into benchmarks.

- No certification activity will be undertaken or endorsed by the standards body: compliance with the ESMF standard will be entirely voluntary.

Next steps

- Formation of an international community standards body, responsible for the maintenance of an open standard for frameworks and meta-data, independent of any implementation.

- PRISM-ESMF Joint Workshop, tentatively planned for 8-9 September 2004, GFDL.

- GO-ESSP workshop, early 2005.