ESMF Strategic Plan for 2012-2015

Last updated: February 28, 2012

The ESMF Mission

- *Earth system models that can be built, assembled, documented, and reconfigured easily, using shared toolkits and standard interfaces.*
- *A growing pool of Earth system modeling components that, through their broad distribution and ability to interoperate, promotes the rapid transfer of knowledge.*
- *Earth system modelers who are able to work more productively, focusing on science rather than technical details.*
- *An Earth system modeling community with cost-effective, shared infrastructure and many new opportunities for scientific collaboration.*
- *Accelerated scientific discovery and improved predictive capability through the social and technical influence of ESMF.*

1 INTRODUCTION

The Earth System Modeling Framework is community-developed software for building and coupling Earth science modeling components. The first phase of the project began in 2002 with a 3-year NASA award to collaborators representing major U.S. modeling centers, and culminated with the delivery of a prototype framework and a set of newly coupled demonstration applications. Also during this period, the NASA Global Modeling and Assimilation Office created the GEOS-5 atmospheric general circulation model, a wholly new model that used ESMF extensively throughout.

The second phase of the ESMF project began as several large programs decided to adopt the framework, and their parent agencies funded the core development team and applications through 2011. New ESMF-based programs and projects included the DoD Battlespace Environments Institute (BEI), the NASA Modeling Analysis and Prediction (MAP) Program for Climate Variability and Change, the NOAA National Environmental Modeling System (NEMS), and a number of smaller projects in related domains such as space weather and sediment modeling. These modeling systems introduced additional requirements, both technically and organizationally. The grid software delivered with the initial ESMF prototype had to be completely redesigned. The composition of the core team changed substantially, in order to bring in staff with greater expertise in mathematics, numerical methods, and application areas.

The management structure of the project was reworked and formalized in order to accommodate multi-agency sponsorship. This organization is described in the ESMF Project Plan[1].

The second phase culminated with the delivery of a production quality, portable, scalable, comprehensive software package, ESMF 5.2.0r.[2] The ESMF software can be used in multiple ways: 1) to create interoperable component-based modeling systems; 2) as a source of remapping, time management, metadata handling, and other functions that can be used independently of the component constructs; and 3) as a fast, file-based offline generator of interpolation weights for many different kinds of grids.

2 CONTEXT FOR THIS PLAN

The ESMF *Strategic Plan 2008-2010*[3] focused on three goals, articulated below:

1. *To complete a stable, high-quality, complete ESMF product, which includes software, documentation, and a support function.* This goal was achieved. For the current *Plan*, the comparable goal is to ensure that ongoing maintenance, user support, and responsiveness to new science requirements and feature requests enable the framework to remain viable.
2. *To develop strategic partners and technological links that enable ESMF applications to operate in increasingly heterogeneous environments and multiple contexts.* For 2008-2010, this meant enabling ESMF components to operate as web services, and creating prototype links to frameworks in other domains such as hydrology. It remains a strategic imperative because it affects how coupled models are constructed for cross-disciplinary applications such as climate impacts modeling, and because the prototypes constructed under the previous *Plan* still need to evolve into production capabilities. An opportunity for use of ESMF in a different context code has appeared with respect to the data analysis and visualization communities, which require many of the advanced grid remapping functions within ESMF, but with a different (python) language binding. Other opportunities for code reuse are likely to emerge.
3. *To support the continued adoption of ESMF by modeling applications.* With about 4000 downloads and 2300 people on its informational mailing list, ESMF is now used in most major climate and weather models in the U.S., and has a growing user base world-wide. Its use has extended to related communities, such as surface modeling. New strategies focus on deepening as well as broadening the use of ESMF, and encouraging codes that are already experimenting with the framework to use more of its features.

---

[1] http://www.earthsystemmodeling.org/management/
[2] http://www.earthsystemmodeling.org/download/ (description as well as download)
[3] http://www.earthsystemmodeling.org/plans/

ESMF has successfully delivered on its early commitments. The state of the framework and its user base, and trends within the larger Earth system modeling community, motivate two new strategic goals.

4. *To build a common model architecture (CMA) that increases interoperability among ESMF applications.* The ESMF-compliant models produced by the first wave of adoption still require manual changes in order to be coupled together. The CMA effort, organized by operational weather centers, introduces additional usage conventions and content standards for ESMF applications, and includes software templates that make it easier for modelers to follow standard architectural patterns. It greatly reduces the number of manual changes that are required to replace a component.
5. *To create self-documenting models and end-to-end workflows.* Increasing scrutiny of model outputs and derived products, the widespread use of ensembles, and the growing complexity of individual models make automated documentation of model provenance especially critical. Increased automation of routine tasks will improve scientific productivity in the face of more complex systems and a distributed, difficult to navigate computational environment.

The strategic goals above are described in more detail in the following section.

3 KEY CHALLENGES, GOALS, AND STRATEGIES

GOAL 1. ONGOING MAINTENANCE, FEATURE REQUESTS, AND USER SUPPORT

The continued use of the ESMF software by the community requires that it be tested and modified to work on constantly changing computing platforms. In order for the software to be highly portable and reliable, the ESMF team tests on about 24 different platforms each night using a growing regression test suite of 4600+ unit tests, system tests, examples, and realistic use test cases. This enables bugs to be identified and fixed shortly after they occur, preventing situations where accumulated problems may mask or confound each other. In software as complex as ESMF – 600,000 lines of code – vigilant maintenance is important.

Optimizations are required as computing architectures shift. Future changes may include updates for codes running on computers with general purpose GPUs or accelerators, where data movement in and out of device memory is the source of the major performance bottleneck. ESMF can help mitigate this performance barrier, especially on the level where it may affect the componentization of model code. There is also potential for leveraging the increased computational power provided by accelerators in the computational intensive portions of the ESMF code, such as the grid remapping algorithms. However, the main focus in ongoing performance optimization of ESMF is to ensure that the software scales readily and does not introduce overhead as codes start to routinely run on hundreds of thousands of processors.

ESMF must be updated for new requirements to remain viable infrastructure for cutting edge science models, and feature requests from users arise frequently even with routine use.

Pending development tasks include the following:

- Additional grid remapping capabilities and performance optimizations. An example is implementation of a 3D conservative grid remapping capability, desired for tightly integrated, high resolution climate-hydrological modeling.
- Better integration of observational data streams into framework data structures and operations such as regridding. Highest priorities are regridding from logically rectangular grids to observational data streams and the ability to construct adjoints of ESMF transformations.
- Fault-tolerant communications.
- Internal interfaces for bringing in external implementations in regridding and I/O.
- Implementation of data operations common to couplers, such as a matrix-vector multiplication.

This is a small sampling of the approximately 180 feature requests that the ESMF project currently has open. In addition, there are about 250 open bugs (note that less than one quarter of these are related to incorrect behavior - other items relate to documentation suggestions, optimizations, refactoring of code, standardization, etc.).

User support remains a vital function of the ESMF team and every day brings new user questions. Although the team strives for excellent documentation, this support function is still critically important to ensuring that users are able to incorporate ESMF correctly into their codes.

GOAL 2. USEOF ESMF IN HETEROGENEOUS ENVIRONMENTS AND MULTIPLE CONTEXTS

The ESMF Strategic Plan for 2008-2010 outlined the challenges of an Earth system modeling community that is really composed of many individual communities: weather forecasting, climate research, hydrological modeling, coastal modeling, surface process modeling, air quality modeling, societial impacts modeling, etc. The models can vary widely in their characteristics and implementation strategies, including coding language, target platform, and level of parallelism. Further, where special services or derivative products are required, the models may be embedded in workflows that contain a variety of other elements, such as a GUI for training students, or a visualization package for creating a set of standard plots. Service oriented architectures offer a universal standard (albeit one that is not appropriate for every software interface) and a connection to ubiquitous web-based environments.

The ESMF team created an option whereby the standard ESMF interface, through a simple switch, can be converted to a web service interface using a Simple Object Access Protocol (SOAP). This interface is being utilized in a pilot effort to connect an atmospheric model

(Community Atmospheric Model) that uses ESMF with a hydrological model (Soil Water Assessment Testbed) that uses the hydrological standard OpenMI. This pilot demonstrates a mode of coupling in which the models from different communities can remain largely intact. The idea is to create interfaces between related communities by connecting the standards used in each of those communities, rather than creating *ad hoc*, one-off systems or requiring that all climate impacts codes be hardwired into a climate model. This approach is still experimental but holds out the promise of preserving local networks built around specialized community models, while still allowing them to interact, and perhaps offer feedbacks, into climate simulations.

ESMF also has the opportunity to repackage some of its functions for other communities. For example, the climate data services community needs parallel grid remapping services like those that ESMF provides. Rather than developing this code again from scratch (a multi-year effort), it would be advantageous to wrap ESMF grid remapping interfaces in the python language so they could easily be accessed by this group and other python users.

The strategic goal here is to successfully carry out the climate-hydrology pilot, and demonstrate the feasibility and advantages of an approach that packages model components as highly fungible objects, capable of spanning multiple contexts and communities. Ideally, other projects will be identified that can reuse the technologies and connections developed through the pilot.

GOAL 3. INCREASED COMMUNITY USE OF THE FRAMEWORK

In the first years of the ESMF project, the outreach emphasis was on encouraging modelers to try the framework. Tutorials focused on how to wrap atmospheric models or other large-scale model components in ESMF interfaces. Many codes were able to do this (the Weather Research and Forecast Model[4], GFDL codes[5], the Community Climate System Model (CCSM)[6], the NOAA Global Forecast System[7], etc.) and to implement simple couplings using ESMF.

Now, the emphasis has shifted to encouraging modelers to use more elements of the framework. Groups such as the Community Earth System Model (formerly CCSM) have found that ESMF can offer high performance, highly flexible grid remapping functions that are not available elsewhere. In particular, the ESMF "offline regrid" parallel application that enables modelers to produce a file of interpolation weights based on two grid files has been extensively used by CESM and other groups. It enables modelers to produce grid files in minutes that would otherwise take hours or might not be possible at all. Good user experiences with this grid remapping function and with other useful libraries, such as calendar management, have resulted in groups that are now curious and receptive to incorporating additional elements of ESMF into their codes.

---

[4] http://wrf-model.org
[5] http://www.gfdl.noaa.gov/fms
[6] http://ccsm.ucar.edu
[7] http://www.emc.ncep.noaa.gov/GFS/

ESMF team outreach could be much more effective, and steps will need to be taken to better organize appearances at conferences and workshops, to produce publications about specific ESMF functions, to offer on-line training materials, and to improve the ESMF website. Future pilot projects may be defined with modeling groups at organizations such as GFDL; emerging science programs such as the NOAA Integrated Water Resources Science and Services (IWRSS) effort and the Earth System Prediction Capability (ESPC); and other infrastructure initiatives such as the European Infrastructure for the European Network for Earth System Modeling (IS-ENES and IS-ENES2)[8].

GOAL 4. A COMMON MODEL ARCHITECTURE FOR INTEROPERABILITY

Interoperability of modeling components can be implemented at a range of different levels. ESMF interfaces enable model components to be called in a standard way, but they do not guarantee that modeling components of the same type (ocean models, for example) can be replaced without further code changes. A key goal for the next five years is to increase the level of interoperability so that fewer code changes are required when model components are exchanged or introduced. This requires action along several fronts. One is promoting a common physical architecture, which defines the scope of components and how they are connected (for example, whether surface processes are handled as a sub-process of the atmosphere). Another is identifying common pieces of code and metadata, standardizing them, and refactoring them into templates that can be reused by multiple groups. Lastly, coupling mechanisms must be structured so that they can be composed and reconfigured more easily, and so that compatibility checking can be automated. These strategies promote a healthy balance between the flexibility needed for research and the constraints required for streamlined operations.

Collectively, these strategies are referred to as "common model architecture" (CMA) by the National Unified Operational Prediction Capability (NUOPC)[9], a consortium of Navy, Air Force and NOAA operational weather prediction centers. This group has worked with a set of researchers drawn from multiple agencies and the ESMF team to define and implement a target level of interoperability for their ESMF codes. The package of templates and conventions being produced is called the NUOPC Layer. The intent is to integrate this Layer, and an earlier interoperability layer developed at NASA called MAPL[10], with each other and the ESMF distribution to the extent possible.

To satisfy this strategic goal, the ESMF team must ensure that the NUOPC Layer is delivered, that target NUOPC operational applications can use it, and that research and other codes begin to work with the software and understand its advantages.

---

[8] https://is.enes.org/

[9] http://www.nws.noaa.gov/nuopc/

[10] https://modelingguru.nasa.gov/docs/DOC-1118

GOAL 5. SELF-DESCRIBING MODELS AND END-TO-END WORKFLOWS

Thorough and accessible documentation of simulations, especially in the climate domain, has become increasingly critical. The ESMF team is committed to enabling models to become "self-describing," and to output metadata along with data, in a format that can be easily ingested by portals that allow it to be searched, browsed, and compared.

Climate model outputs are public records which policy makers and resource managers look to for guidance on the impacts of climate change. Documentation is needed for understanding and using the records, distinguishing between runs, and establishing their credibility. While it may have been sufficient in the past to track information about a simulation back to a set of scientific papers and perhaps acquire the rest through personal inquiries, the practical use of climate data (and to a lesser extent, active models) by a broad swath of applications requires that information be more complete and accessible. The growing complexity of Earth system models and the use of ensembles for evaluation and prediction make the need more urgent. With hundreds of thousands of lines of code and countless run time and post-processing options, it is difficult for modelers to document simulations manually. Understanding the configuration of an Earth system model and comparing it to others demands that models be associated with structured metadata and that to the extent possible, this information is generated and archived automatically with model runs.

The ESMF team has been working with the European METAFOR[11] consortium to create a Common Information Model (CIM) for Earth system models, and to use the CIM in a variety of different efforts. The CIM itself builds upon and extends widely used community standards, including ISO and Climate and Forecast Conventions, and is the metadata being used in the 5th Coupled Model Intercomparison Project (CMIP5)[12], an integral part of the next Intergovernmental Panel on Climate Change (IPCC) Assessment. For CMIP5, modelers were required to fill out a questionnaire to describe their model, a process that was arduous, time consuming, and would need to be repeated for new experiments. The ESMF team saw an opportunity here to improve efficiency by enabling the framework to produce much of this metadata automatically.

ESMF includes a class called Attributes that enables comprehensive metadata to be written out from models in the form of a CIM-compliant XML file. The ESMF team implemented this capability in the Community Earth System Model (CESM) in collaboration with NCAR model developers. Together with other collaborators from Purdue and the Earth System Grid, this team created an automated workflow that enables the CESM model to be configured and run through a GUI, with the output data and CIM metadata automatically output and then ingested by an Earth System Grid Gateway[13]. The metadata is searchable, browsable, and linked directly to the data

---

[11] http://metaforclimate.eu
[12] http://cmip-pcmdi.llnl.gov/cmip5/
[13] http://earthsystemgrid.org

from the run, making the model output far more usable. In the future the ESMF team does not anticipate developing GUIs or data distribution systems, but does foresee additional collaborations of this type in order to develop other end-to-end, documented workflows.

Self-documentation is not the only advantage of such end-to-end workflows. The GUI interface and automated processing steps increase ease of use, and make models more accessible to new users and students.

It is a strategic goal to evolve self-describing workflows from prototype to production quality, to encourage other models to try the capabilities, and to include more aspects of provenance metadata. In particular, a gap currently exists in the ability to automatically generate and represent in a standard way post-processing steps applied to the data, a critical omission. The ESMF team will work with collaborators to address this gap.

4 STRATEGIC VISION

By the close of 2015, ESMF will be recognized as an essential element of the infrastructure used in U.S. Earth system modeling, and will be used in all major U.S. climate and weather models. The base software will be updated for evolving computational platforms and user needs. The NUOPC Layer will be in operations at the National Weather Service and Navy models, and its use will be starting in emerging efforts such as the Integrated Water Resources Science and Services (IWRSS) and Earth System Prediction Capability (ESPC) programs. More than 100 ESMF weather/climate/water components will be available nationwide, and the framework will offer service interfaces that enable some of these components to couple to those from other disciplines, to be incorporated into complex workflows, and to be reused by communities such as climate data services. A portfolio of instructional and informational materials will be available, including on-line examples and tutorials that enable individuals to learn to use the software without significant assistance from the ESMF support team. Several modeling groups will be using ESMF to produce self-documenting simulations within end-to-end workflows, including documentation of post-processing steps. ESMF will have an established outreach and training program, and will continue to be governed by a multi-agency consortium.