ESMF Strategic Plan for 2008-2010

February 25, 2008

The ESMF Mission

- *Earth system models that can be built, assembled 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.*

Introduction

ESMF is community-developed software for building and coupling Earth science modeling components. The project began in 2002 with a 3-year NASA award that funded a core development team and a number of groups with modeling and data assimilation applications. The first phase of the project culminated with the delivery of a prototype framework and a set of newly coupled demonstration applications. In addition, the NASA Global Modeling and Assimilation Office created the GEOS-5 atmospheric GCM, 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 team and applications through 2010. ESMF-based programs and projects funded included the DoD Battlespace Environments Institute, the NASA Modeling Analysis and Prediction Program for Climate Variability and Change, the NOAA National Environmental Modeling System, 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].

1. Key Challenges, Goals, and Strategies

ESMF is now three years into its second funding cycle. The project must prepare for the completion of the cycle, assess outstanding community needs, and envision what comes next. Computing and science requirements have evolved. For example, the recognition of global warming as a credible and manifest phenomenon has shifted the nature of the climate science questions emerging from the policy community, national and regional resource management agencies, businesses, and the public. Rather than focusing on establishing scientific consensus on the reality of global warming, current interest focuses on adaptation and mitigation. The models, datasets, methodologies, and research communities involved in these endeavors can be quite different from those involved in traditional global climate modeling. Assessments of regional impacts may require models of specific local-scale processes, such as wildfires and storm surges. Climate or weather models may be used simply as boundary conditions. The local-scale models are often written for Windows platforms instead of supercomputers, in languages other than Fortran, and may have products, such as impact maps, that are destined more for analysis by regional managers than by scientists at a national laboratory or university. The modeling infrastructure for these applications must transition easily between desktop systems and high performance computing systems and must enable heterogeneous combinations of data services, models, and visualization and analysis.

Technology trends also point to more heterogeneous environments with loosely coupled[2] elements. There is growing hardware support for this via the creation of high-bandwidth networks that connect computing resources. Low-level software utilities, such as Globus, that enable networks to be linked into regional and national computing grids are now second generation and increasingly robust. Services can be deployed and interact over networks using web service protocols, a form of loose coupling. To help users discover and access resources in these environments, projects such as the Earth System Grid[3], the Global Change Master Directory[4], and the Community Data Portal[5] are building integrated catalogs of models, datasets, and tools. In some cases these "science gateways" use workflow packages such as Kepler[6] to create and reproduce patterns of interaction.

At the other end of the computational spectrum, researchers seek to refine their knowledge of global circulations and weather phenomena through increasingly high resolution models. These models may require running on petascale supercomputing platforms using thousands or tens of thousands of processors. Their infrastructure must scale out to these processor counts in memory

[1] http://www.earthsystemmodeling.org/management/
[2] By tight coupling we mean interactions that require the sort of frequent, high volume, and rapid data communications you might find between, say, an atmospheric model and a land model. By loose coupling we mean those interactions that do not.
[3] http://www.earthsystemgrid.org
[4] http://gcmd.nasa.gov
[5] http://cdp.ucar.edu
[6] http://kepler-project.org

and performance, a challenge that sometimes means completely redesigning software originally designed for tens or hundreds of processors.

The emerging petascale requirements fall into the realm of business as usual, albeit a challenging business, for ESMF. Continuing performance studies[7] indicate that ESMF can operate efficiently at very high processor counts, and that the modifications to its source code needed to improve it further are non-systemic.

The real challenge for ESMF is not in adapting for petascale platforms, but in bridging the gap between high performance computing and desktop computing, and enabling ESMF components to take advantage of the heterogeneous environments necessary for greater synthesis and information flow across research, operational, military, commercial, and governmental enterprises. ESMF cannot do that independently. Due to the range of requirements it must meet and the number of partners it involves, the project is already technically and organizationally challenging. Instead, the strategy for ESMF will be to limit the project scope, complete the items within that scope, and define a strategic position and partners that enable it to function as a key element of an evolving, interlocking, international network of science support tools.

It will be most cost efficient for the ESMF project to take advantage of existing paradigms, such as service oriented architectures, and projects that have built up general toolkits for heterogeneous environments, such as CCA's Babel[8], a computer language translator. ESMF applications may need to interact with frameworks from other domains, such as OpenMI[9], a hydrology framework written in C# for Windows. The ideal role for ESMF – one that protects investments, and encourages rather than hinders flexibility - is to leverage the interface and metadata standardization implicit in adopting the framework to enable ESMF-compliant components to operate in both traditional and heterogeneous environments. We are part of the way there. Interface standardization follows naturally from the existing ESMF implementation. Metadata descriptions are an area of active development, in conjunction with the Curator project[10], the Earth System Grid, and the European METAFOR project[11]. The translation of ESMF components into services that can be loosely linked requires additional definition and possibly additional funding.

This strategic plan thus addresses three overriding goals:

1. To complete the requirements specified for ESMF at its inception, and the advanced functional requirements that follow naturally from them. The initial requirements were for infrastructure and coupling tools for high performance, tightly coupled applications in climate, weather, and related disciplines; advanced requirements include petascale performance, and more sophisticated (unstructured, multi-tile, nested) grids. Priorities for functional requirements will be set by the ESMF Change Review Board.

---

[7] http://www.earthsystemmodeling.org/metrics/performance/
[8] http://www.cca-forum.org
[9] http://www.openmi.org
[10] http://www.earthsystemcurator.org
[11] http://www.prism.enes.org

2. To develop strategic partners and technological links that enable ESMF applications to operate in increasingly heterogeneous environments.

3. To support the continued adoption of ESMF by modeling applications.

2. Strategic Vision

By the close of 2010, ESMF will be the established methodology for coupling Earth system modeling applications at a critical mass of national and international laboratories and universities. It will be used by research centers including NASA GSFC, DoD NRL, NOAA GFDL, and NCAR as well as operational centers such as NOAA NCEP, FNMOC, NAVO, and AFWA. The software will enable models from different groups to be coupled together with less effort than currently required, and with more flexibility in configuration and numerical approach. More than 100 ESMF components will be available nationwide. The software will be stable, high-performance, portable, straightforward to learn and use, and consistent in interface and implementation. It will satisfy the coupling requirements of tightly coupled weather, climate, and related applications. 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. It will be simple to test components for ESMF compliance. Presentations and publications will be available that describe the evolution of the project, its technical strategies, use in applications and resulting scientific achievements, and lessons learned. The project will have an ongoing governance and outreach program and a maintenance plan that extends beyond 2010. The ESMF project will develop strategic partnerships and technical compatibilities that enable it to evolve with the changing computational environment, to be used with key emergent technologies, and to adapt to new scientific requirements.

3. Strategic Goals and Supporting Goals

Goal 1: Produce a stable, high-quality, complete ESMF product. The primary focus of the ESMF team over the next three years will be the completion of the ESMF product, which includes software, documentation, and a support function.

Supporting Goal 1a: Complete functional requirements. There are outstanding pieces of development that must be delivered to support core functions. These are: 1) complete an ArrayBundle class than enables optimized data transfers of multiple Arrays, 2) complete the Field and Bundle communications methods, 3) integrate contributed location stream code for the representation of observational data, 4) complete arbitrary distributions for structured Grids, introduce multi-tile Grid methods, and implement exchange Grids, 5) complete the interface to a general regridding package, 6) complete the interface to unstructured meshes, 7) complete a C interface, 8) complete a hierarchical attribute package, 9) complete a rudimentary IO package, 10) complete a compliance tester for components. The ESMF team must also address the difficulties in obtaining adjoints caused by component indirection, and address build improvements needed for looser coupling (see Goal 2a).

Measures: Schedule and complete the necessary functionality.
Metrics: Releases containing new capabilities and use of the capabilities in applications.
Time frame for completion: Items 1-8 completed by the end of calendar year 2009. The remaining items should be completed by end of calendar year 2010.

Supporting Goal 1b: Demonstrate performance portability on a range of platforms. To date ESMF has demonstrated low overhead and excellent performance up to 2000 processors, which is the maximum that has been available for testing. There are new challenges posed by platforms with 10,000+ processors. ESMF will continue to evaluate and optimize performance and memory use to ensure that it is comparable or superior to native implementations and other packages that perform similar functions. In order to ensure portability, ESMF will continue its exhaustive regression testing program. Further, ESMF will continue to explore improved testing strategies, including implementation of a harness for testing the large parameter spaces of grid-dependent methods, and use of the TeraGrid for accessing new platforms and acquiring additional computing time.

Measures: Complete studies that evaluate ESMF performance on a range of processor counts and for a variety of configurations. Regression test ESMF regularly on a wide variety of platforms.
Metrics: Number of platforms tested and supported. Overhead and scalability measurement studies show comparative (target < 5% overhead) or superior performance relative to software with similar functionality.
Time frame for completion: Ongoing with development of new software functions.

Supporting Goal 1c: Improve the ease of learning and using ESMF. One of the challenges of the ESMF project is supporting a large set of functions and options in the software while still ensuring that it is easy to understand and use. Adopting ESMF can still be daunting, so this is a critical focus area. In order to achieve ease of use, ESMF must 1) resolve a backlog of bug reports, support requests, and feature requests, 2) standardize software interfaces and implementation for consistent look and feel and behavior, 3) expand the portfolio of instructional materials and improve the tutorial program overall.

Measures: Resolve backlog of bug reports, support requests, and feature requests. Continue standardization of interfaces and implementation. Enhance the portfolio of instructional materials, including development of on-line tutorials.
Metrics: Number of open tickets. Releases with greater standardization across interfaces and behavior. Increased user satisfaction with instructional materials.
Time frame for completion: Completed by 2010.

Supporting Goal 1d: Attract new users and create a positive impression of ESMF in the broader community. The ESMF project had an intense period of publicity at its inception, and another with the delivery of demonstration applications at the end of the first funding cycle. The ESMF team plans one more round of publicity to coincide with the delivery of a functionally

complete package and a paper reviewing the transformational impact of ESMF, technically and socially, across the modeling community. In the interim, the team plans to continue journal publications, presentations at technical conferences, visits to customer sites, community meetings, tutorials and coding camps, seminars, and other outreach efforts. In order to coordinate these efforts and build to a period of more focused attention on the project an Outreach Plan will be developed.
Measures: Continue presentations, publications, and tutorials. Develop an Outreach Plan.
Metrics: Number of journal publications, number of presentations, number of training classes and seminars, number of individuals trained. Awareness and positive impressions of ESMF throughout the broader community.
Time frame for completion: Outreach Plan completed during 2008 and implemented thereafter. Review paper on ESMF impacts by 2010. Topical journal papers and conference presentations on lessons learned, technical achievements, process/governance, etc. as opportunities arise. At least one user training session per year.

Goal 2: Prepare ESMF systems for integration into increasingly heterogeneous computing environments. ESMF initially focused on coupling components intended to run on the same computer, with performance as the foremost concern. In response to changing science requirements and technical trends, we will leverage the interface and metadata standardization implicit in ESMF adoption in order to enable ESMF components to operate in more heterogeneous environments. One aspect of this is linking ESMF components to loose coupling technologies. Another is introducing ESMF components and models into "science gateways" that catalog and integrate diverse, distributed resources.

Supporting Goal 2a: Link ESMF components to loose coupling technologies. Explore technologies for serializing ESMF objects and wrapping them with web service interfaces. Modify the ESMF build so that it is easier to compile components into separate shared objects and executables. Define interfaces between ESMF and middleware projects and frameworks, such as OpenMI, CCA, and InterComm, that support looser coupling. Work with modeling groups interested in using these technical strategies to create coupled systems.

Measures: Implement links to looser coupling strategies as time and existing resources permit. Pursue new funding opportunities that would enable further development in this area.
Metrics: ESMF releases with extensions for looser coupling. Demonstration applications featuring alternative forms of coupling. Funds for new development.
Time frame for completion: Proof of concept by 2009.

Supporting Goal 2b: Integrate ESMF components and models into science gateways. Continue collaborative development of model metadata and work with the Earth System Grid, Community Data Portal, and other efforts to incorporate model metadata into their ontologies. Introduce ESMF itself and ESMF components and models into science gateways.

Measures: Much of this work will be completed as part of the Earth System Curator project. Curator is linked to ESMF through shared leads DeLuca (NCAR) and Balaji (GFDL).
Metrics: Installation of ESMF and compliant components and models in science gateways.
Time frame for completion: Prototype during 2008.

Goal 3: Research and operational codes nationally and internationally adopt ESMF. ESMF initially focused on adoption by major research and operational centers, for a number of reasons. First, national centers acting in unison to adopt ESMF sent a strong message to the rest of the community, and represented a manageable subset of geoscience applications. Second, since the national centers serve an integrative function and support large multi-component codes, they have more of a need for coupling than many of the smaller university programs. Third, the major sponsors of ESMF (DoD, NASA, NOAA) have provided more funds and application support for adoption at national centers than within the university community. ESMF will continue to focus on adoption support efforts at major centers, extending outreach to the university community as time and resources allow.

Measures: Increase each year the number of modeling components and models using ESMF.
Metrics: Number of ESMF components available. Number of ESMF applications in routine use or operations.
Time frame for completion: Goal is 100 components available by 2010, 20 major modeling applications in routine use or operations by 2010.

4. Management and Implementation

The Strategic Plan was reviewed and approved by the ESMF Executive Board. It was circulated for comment to the ESMF Advisory Board and members of the Interagency Working Group. The plan will be carried out by the ESMF Core Team, Change Review Board, and other project bodies described in the ESMF Project Plan.

Annual reports will be prepared at fiscal year end that summarize actual performance of the project during the year, and describe specific goals, metrics, and implementation strategies for the coming year. These reports will be circulated to ESMF Executive management for review and will be posted in the Work Plans section of the ESMF website.[12]

[12] http://www.earthsystemmodeling.org/plans/annual.shtml