



Jun 16th, 10:40 AM - 12:00 PM

NESII Modular, High Performance Infrastructure for Earth System Modeling

Cecelia Deluca

Cooperative Institute for Research in Environmental Sciences, cecelia.deluca@noaa.gov

Peggy li

California Institute of Technology

Gerhard Theurich

Science Applications International Corporation

Follow this and additional works at: <https://scholarsarchive.byu.edu/iemssconference>

 Part of the [Civil Engineering Commons](#), [Data Storage Systems Commons](#), [Environmental Engineering Commons](#), [Hydraulic Engineering Commons](#), and the [Other Civil and Environmental Engineering Commons](#)

Deluca, Cecelia; li, Peggy; and Theurich, Gerhard, "NESII Modular, High Performance Infrastructure for Earth System Modeling" (2014). *International Congress on Environmental Modelling and Software*. 7.

<https://scholarsarchive.byu.edu/iemssconference/2014/Stream-B/7>

This Event is brought to you for free and open access by the Civil and Environmental Engineering at BYU ScholarsArchive. It has been accepted for inclusion in International Congress on Environmental Modelling and Software by an authorized administrator of BYU ScholarsArchive. For more information, please contact scholarsarchive@byu.edu, ellen_amatangelo@byu.edu.

NESII Modular, High Performance Infrastructure for Earth System Modeling

Cecelia DeLuca¹, Peggy Li², Gerhard Theurich³

¹NOAA Environmental Software Infrastructure and Interoperability Group,
Cooperative Institute for Research in Environmental Sciences, 325 Broadway St,
Boulder, CO 80303, cecelia.deluca@noaa.gov

²Jet Propulsion Laboratory of the California Institute of Technology,
4800 Oak Grove Drive, Pasadena, CA 91109

³Science Applications International Corporation,
1710 SAIC Drive, MacLean, VA 22102

Abstract: The mission of the NOAA Environmental Software Infrastructure and Interoperability (NESII) group is to advance understanding and improve predictions of the Earth system by delivering infrastructure software that enables new scientific discoveries, fosters collaborations, and promotes resource efficiency. NESII software supports environmental modeling, data search and analysis, and metadata services. A core capability is the Earth System Modeling Framework (ESMF), a comprehensive object-oriented framework for building and coupling models. However, new development has focused on modular packages that can easily be combined with other software elements to form solutions to complex problems. Such modular packages include ESMPy, a Python grid remapping utility; OpenClimateGIS, a Python toolkit for subsetting, reformatting, and performing computations on climate data; Earth System CoG, a collaboration environment focused on model intercomparison projects; and Earth System Documentation (ES-Doc) tools for collecting, displaying, and comparing model metadata, developed in collaboration with an E.U. consortium. These and other NESII efforts are described.

Keywords: high performance computing, software infrastructure, model coupling, framework, metadata, Python

1. OVERVIEW

The NESII group (<http://www.esrl.noaa.gov/nesii/>) collaboratively develops a range of software infrastructure products for the Earth system sciences. These include model coupling systems, grid remapping and other utilities, metadata services, data subsetting and reformatting tools, and model intercomparison and collaboration environments. NESII products are listed in Table 1. They can stand alone, but they are also built to work together as a suite to address complex problems.

The core scientific problem that motivates NESII software development is modeling the Earth system in order to make predictions about weather and climate. This problem has scientific, technical, social, and semantic aspects, which are often interlinked. It transcends the efforts of a particular organization, agency, or domain. NESII customers include modeling groups from universities, the Navy, the National Center for Atmospheric Research, the National Weather Service, the Department of Defense, and NASA. These organizations, who develop models of the atmosphere, space weather, ocean, wave, land, ice, and other natural systems, must work together to combine them into composite models, and to assess their results.

NESII efforts include:

- The Earth System Modeling Framework (ESMF)¹, a software architecture for composing complex, coupled modeling systems

- The National Unified Operational Prediction Capability (NUOPC), a layer that introduces conventions and templates for ESMF-based models in order to increase their interoperability
- The ESMF Python interface (ESMPy), a Python interface to the ESMF high performance regridding utility
- The Earth System CoG (CoG), a collaboration environment that combines project hosting and networks, metadata and data services, and support for distributed governance
- Earth System Documentation(ES-Doc), an international effort to develop metadata services for a set of climate modeling and related projects
- ESMF Web Services, tools for building distributed modeling systems that can incorporate both high performance computing and local, PC-based computing resources
- Open Climate GIS (OCGIS), a Python package designed for geospatial manipulation, subsetting, and computation of climate datasets stored in NetCDF files

In the following sections, we describe NESII tools and the types of applications they have been used to address.

2. RELATED WORK

Many efforts to develop coupling frameworks and modeling infrastructure exist in other regions, such as the European OASIS coupler, and in other domains, such as the Community Surface Dynamics Modeling System (CSDMS). A useful review of related technologies is provided in Valcke et al. 2012.²

3. NESII SOFTWARE

Weather and climate models, and the infrastructure that they use, must be able to effectively exploit high performance computer resources. The NESII team tracks technical trends and optimizes its software for current and emerging supercomputer architectures. The team regularly assesses the performance of its software relative to custom infrastructure solutions, and maintains a page of performance reports.³

However, an efficient parallel implementation and ongoing performance optimization is just one aspect of the software development required to build and deliver high quality modeling and data systems. It is also essential to deliver

software that is accessible, thoroughly tested, documented, and supported. These are important attributes of NESII software and they affect the role it is able to play in the modeling community. For

Product	Description
ESMF	High performance model coupling and toolkits http://www.earthsystemmodeling.org/
ESMF_RegridWeightGen	Parallel, file based utility for generating interpolation weights http://www.earthsystemcog.org/projects/regridweightgen/
ESMF Python (ESMPy)	Python interface to ESMF grid remapping http://www.earthsystemcog.org/projects/esmpy/
NUOPC Layer	Templates and conventions for increased model interoperability http://www.earthsystemcog.org/projects/nuopc/
ESMF Web Services	Tools for operating and linking ESMF components as web services http://www.earthsystemcog.org/projects/esmfwebservices/
Earth System CoG	Project workspaces with linked data services http://www.earthsystemcog.org/projects/cog/
ES-DOC	Model metadata form generation tool and other metadata tools and standards (an E.U.-led effort) http://earthsystemcog.org/projects/es-doc-models/
OpenClimateGIS	Tools for subsetting, reformatting, and calculating derivatives of climate data http://earthsystemcog.org/projects/openclimategis/

Table 1: A summary of NESII products.

example, the ESMF package includes 6000+ regression tests that run nightly on 30+ platform/compiler combinations daily, in addition to comprehensive documentation and a support team. This reliability enables NESII products to be used in mission critical, operational settings, and to serve as a link between these applications and research applications. The products and development strategies selected by NESII are always tied to the goal to create open-source software products that deliver production-ready capabilities that are not available elsewhere, and enable the solution of problems that would otherwise be intractable.

2.1 Earth System Modeling Framework (ESMF)

The **Earth System Modeling Framework (ESMF)** was developed by NESII in concert with partners from NOAA, NASA, DoD, DOE, and NSF centers and universities. ESMF software enables modelers to wrap components with standard interfaces, and transform and transfer data between components while a coupled model runs. It also includes fully featured, portable, high performance parallel libraries for common modeling functions: calendar management, grid remapping and data communication, I/O, and automatic generation of model metadata. Many U.S. models used for weather forecasts, coastal modeling and predictions, and climate projections rely on ESMF to integrate components from multiple sources and create flexible, reconfigurable systems.

There are over 85 ESMF components and 12 ESMF-based multi-component modeling systems, in weather, climate, hydrology, and space weather. For a partial list, see: <http://www.earthsystemmodeling.org/components/>

2.2 Grid Remapping with ESMF_RegridWeightGen and ESMP

One of the most powerful aspects of ESMF is the range of data structures that it accommodates: unstructured meshes, logically rectangular grids, observational data streams, and “exchange grids” – merged grids used for conservative exchange of fluxes. ESMF supports fast, parallel remapping among these different data structures. There are options for remapping in 2D and 3D, with regional and global grids, and with a variety of interpolation methods (higher order, bilinear, first order conservative, nearest neighbor).

A very easy way to gain access to ESMF grid remapping is a utility called **ESMF_RegridWeightGen**, which ingests grid files, and, in parallel, calculates and outputs interpolation weights. This software quickly found users after the 5th Coupled Model Intercomparison Project (CMIP5)⁴ data was collected. The climate grids in CMIP5 models were less structured and higher resolution than ever before. The remapping tools in analysis and visualization packages such as the NCAR Command Language (NCL)⁵ could not handle their size and complexity, and adding ESMF remapping allowed scientists to process outputs from models that they could not handle previously.

ESMPy, a thin Python wrapper around ESMF grid functions, was designed to support the many Python users in the Earth system data services community. These users sought a way to bring ESMF grid remapping into their data analysis scripts. Like ESMF_RegridWeightGen, it was quickly picked up by individuals and software packages, including NOAA’s Ferret⁶ and the DOE-funded Ultrascale Visualization and Data Analysis Toolkit (UV-CDAT)⁷ to process next-generation, highly scalable climate grids.

2.3 The NUOPC Layer

The National Unified Operational Prediction Capability (NUOPC)⁸ is a consortium of NOAA and Navy operational weather prediction centers and their research partners. The consortium is creating a set of software templates, conventions, and tools that increase interoperability of ESMF-based modeling applications. This **NUOPC Layer**, based on ESMF, helps to ensure that even components with complicated coupling scenarios – implicit and semi-implicit coupling, nested components, ensembles – fit together properly. The NUOPC Layer has been introduced into Navy regional and global coupled modeling systems used for ocean prediction, the NOAA NCEP codes used for weather forecasting by the National Weather Service, and the Community Earth System Model (CESM), which is used for climate research and national and international assessments. Work is also in progress to introduce the NUOPC Layer into the NASA Model E code and the GEOS-5 atmospheric general circulation

model. The integration of the NUOPC Layer into these codes standardizes the implementation of ESMF so that they can more easily contribute components to jointly developed systems, and exchange components with each other.

2.4 ESMF Web Services

Web services are a way to utilize scattered resources, collect and deliver information, and manage transactions across computing networks that may be distributed, heterogeneous, and vast. They offer effortless access; anyone who reads a blog or makes a purchase on the internet relies on web protocols. Traditional high performance computing, on the other hand, typically requires highly trained users, who run arcane code on masses of tightly interconnected processors behind firewalls. The NESII team saw an opportunity to extend ESMF in order to bridge the gap between these worlds, and address an emerging issue relevant to NOAA and other agencies: how to integrate projections from high performance climate models with the local tools used by practitioners of water management.

The NESII team and collaborators assembled a proof-of-concept two-way coupled system that linked the Community Atmosphere Model⁹ (CAM, running on a high performance system at Oak Ridge) with the Soil and Water Assessment Tool¹⁰ (SWAT, running on a personal computer), using web services to transfer data.¹¹ This prototype is of interest to resource managers and others who are looking for ways to include information about the changing climate in their products, while preserving local information delivery systems.

The coupled hydro-climate code relies on an **ESMF Web Service** switch that enables a component with an ESMF wrapper to run as a web service. This web service option is included in the ESMF distribution, and the coupled system is also available for download. New work will update the coupled system so that it can use the Weather Research and Forecast Model (WRF)¹² as well as CAM.

2.5 Earth System CoG

ESMF, NUOPC, and other NESII tools enable modelers to exchange components, perform controlled experiments, and engage in comparative analyses. Building a software environment that supports collaborations and comparisons was a natural progression for the NESII team.

Earth System CoG is a NESII-developed web environment that enables users to create project workspaces, connect projects into networks, share and consolidate information within those networks, and seamlessly link to tools for data archival, reformatting and search, data visualization, and metadata collection and display. CoG is integrated with the Earth System Grid Federation (ESGF)¹³ data distribution software and provides an easy to use interface to its services. While other collaboration environments exist, CoG's unique combination of capabilities is ideal for model intercomparisons, ensemble analysis, and other multi-site, data-centric endeavors.

CoG demonstrated its value during the *Dynamical Core Model Intercomparison Project (DCMIP)*¹⁴, a two week colloquium held at NCAR during 2012. Participants performed a set of objective tests on a selection of dynamical cores – the part of a model that solves the fluid equations – developed by modelers from around the world. CoG enabled each participating group to maintain a separate workspace, as well as common spaces where results could be archived, searched, visualized, and discussed. Figure 1 offers a quick look at the DCMIP home page on the CoG website.

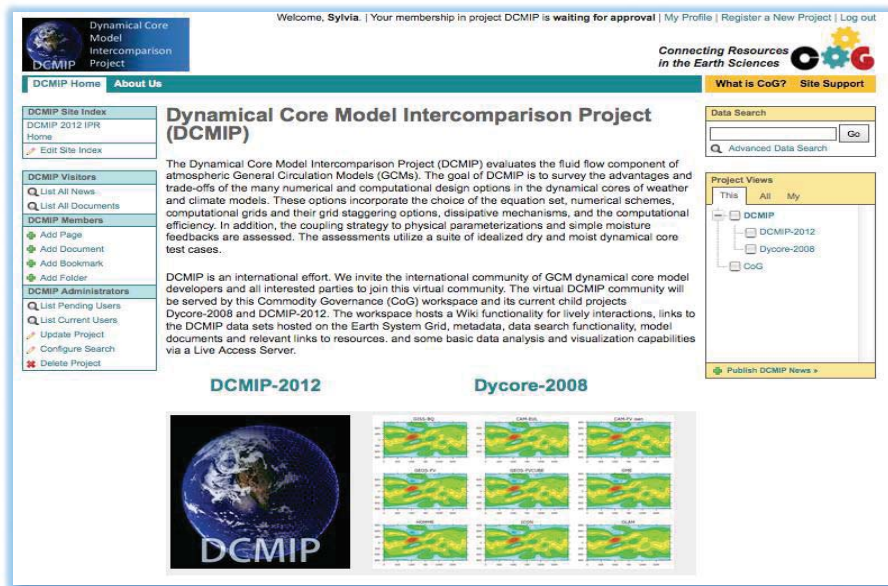


Figure 1. Screen shot of the DCMIP project page on Earth System CoG.

In all, about 75 projects are now using CoG, including the National Climate Predictions and Projections (NCP) platform, a suite of projects focused on providing local and regional information about climate, and the High Impact Weather Prediction Project (HIWPP), which is focused on exploring options for the next operational weather prediction system.

2.6 ES-DOC

Model ensembles and collective analysis became matters of public interest with the IPCC Fourth Assessment Report (AR4)¹⁵. Much of the analysis for that report came from the 3rd Coupled Model Intercomparison Project, CMIP3¹⁶, which compared runs from a collection of different climate models configured for specific scenarios. Although AR4 had tremendous impact and influence, afterwards scientists noted ways to improve the process for the next assessment, AR5. Collecting additional metadata describing the models emerged as a high priority, since limited metadata was collected for CMIP3 and it could be difficult to find relevant information about the model runs once they were completed. Preparations for AR5 motivated new projects in the E.U. (the METAFOR project¹⁷) and the U.S. (the Earth System Curator project¹⁸, led by the NESII team) focused on metadata for assessments. These efforts created several prototype products while U.S. and E.U. participants became increasingly familiar with the others' work.

The **Earth System Documentation (ES-Doc) project** merges and replaces METAFOR and Earth System Curator. It is an international collaboration, led by the E.U., that is continuing to evolve climate metadata standards and tools. Its participants worked together to develop the Common Information Model (CIM) schema for the description of climate models, and tools for collection and display of metadata. These tools were deployed for the 5th Coupled Model Intercomparison Project (CMIP5), on which AR5 will be based.¹⁹ The CIM includes comprehensive descriptions of models, simulations and experiments and will vastly increase the amount of model information available and accessible for AR5.

While a primary motivation for ES-Doc remains IPCC and other large-scale assessments, the metadata tools being created by this collaboration matter to NESII and the broader community for other reasons. Such metadata tools are useful whenever scientists need to construct, describe, or justify a multi-model ensemble, run a component or model intercomparison project, or perform collaborative data analysis; that is, all the time. NESII's main contribution to the ES-DOC tool set has been a form generator that easily creates a CIM-based web form for metadata input. E.U partners are creating updated versions of CIM viewers and comparators.

In addition to their use in CMIP5, ES-Doc tools were integrated into the CoG environment in order to support the DCMIP project, and were used in a 2013 downscaling workshop hosted by the National Climate Predictions and Projections Platform (NCPP)²⁰.

2.7 OpenClimateGIS

OpenClimateGIS software provides access to climate model projections and climate derivatives in commonly used, modern formats used by GIS software, browser-based mapping tools, and virtual globes. The OpenClimateGIS system does not store climate data archives locally, but rather works in conjunction with external climate archives that expose climate data via the OPeNDAP²¹ protocol. In addition to being a “stand-alone” software program that can be used by other software or inside analysis scripts, OpenClimateGIS can operate as a web service accessible via a web browser or other web clients.

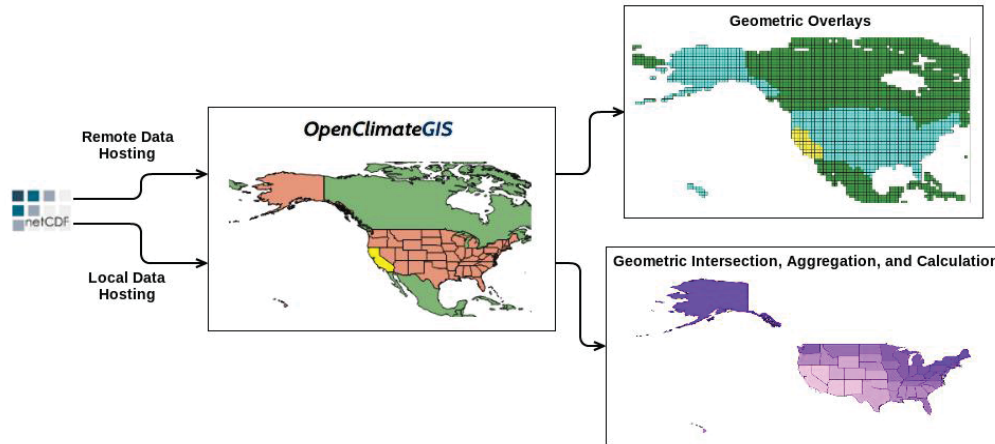


Figure 2. OpenClimateGIS reads, processes, and reformats climate data from local or remotely hosted climate data archives. Gridded climate data is subsetting using a combination of arbitrary geographic boundaries (e.g. watershed), latitude/longitude bounding box, time range, and/or level range.

OpenClimateGIS offers a convenient way to convert native climate model formats such as NetCDF²² to data formats associated with geographic analysis. For example, in addition to a bounding box, OpenClimateGIS can subset a climate dataset following arbitrary boundary definitions such as a watershed, county, or country. Supported geometric operations include intersects/overlay and intersection/clip with the option to aggregate spatially coincident climate data to its corresponding geometry via area-weighting. A user can apply a variety of computations to a subsetting data “cube.” OpenClimateGIS then writes the data subset to common formats easily ingested by GIS software (e.g. ESRI ArcGIS²³, or the open source QGIS²⁴) or statistical software such as R²⁵. Figure 2 is a simple schematic of OpenClimateGIS capabilities.

NESII Sponsors and Collaborators



4. SOFTWARE GOVERNANCE

NESII projects are managed as multi-agency, distributed, community-driven efforts. Users set development priorities, and review design and implementation strategies. Software is distributed under open source licenses and, to the extent possible, processes are open and project information is stored publicly. NESII development is distributed and the team is committed to supporting an international and diverse set of developers and customers.

5. CONCLUSION

All of the products described in this document are related to some degree, through reuse of a core set of software technologies: ESMF, ESGF nodes, and the CIM schema. These elements took years to collaboratively develop, test, and refine, led by NESII, in the case of ESMF, and its collaborators, for ESGF and CIM. NESII has leveraged these community investments by assessing needs and adapting, wrapping, reconfiguring, connecting, and otherwise transforming these technologies to quickly create derived, modular products such as ESMF_RegridWeightGen and ESMPy. The NESII team has also established complementary efforts such as OpenClimateGIS, which can be used alone or combined with other core technologies within multi-purpose environments such as CoG. By building on core technologies in modeling (ESMF), metadata (ES-Doc/CIM) and data services (ESGF/CoG), the NESII team has been able to address a range of community needs.

5. REFERENCES

-
- ¹ Hill, C., C. DeLuca, V. Balaji, M. Suarez, and A. da Silva (2004). Architecture of the Earth System Modeling Framework. *Computing in Science and Engineering*, Volume 6, Number 1, pp. 18-28.
 - ² Valcke' S., V. Balaji, A. Craig, C. DeLuca, R. Dunlap, R. W. Ford, R. Jacob, J. Larson, R. O'Kuinghttons, G. D. Riley, and M. Vertenstein' (2012). Coupling technologies for Earth System Modelling, *Geosci. Model Dev.*, Volume 5, pp. 1589–1596.
 - ³ ESMF performance reports, <http://www.earthsystemmodeling.org/metrics/performance/>
 - ⁴ CMIP5 overview, <http://cmip-pcmdi.llnl.gov/cmip5/>
 - ⁵ NCL homepage, <http://www.ncl.ucar.edu/>
 - ⁶ Ferret homepage, <http://ferret.pmel.noaa.gov/Ferret/>
 - ⁷ UV-CDAT homepage, <http://uv-cdat.llnl.gov/>
 - ⁸ NUOPC Content Standards Committee, NUOPC Layer Reference Manual, http://www.earthsystemmodeling.org/nuopc/NUOPC_refdoc/
 - ⁹ CAM homepage, <http://www.cesm.ucar.edu/models/cesm1.1/cam/>
 - ¹⁰ SWAT homepage, <http://swat.tamu.edu/>
 - ¹¹ Goodall, J. L., K. D. Saint, M. B. Ercan, L. J. Briley, S. Murphy, H. You, C. DeLuca, R. B. Rood (2013). Coupling climate and hydrological models: interoperability through web services, *Env. Modeling and Software*, Volume 46, pp. 450-459.
 - ¹² WRF homepage, <http://www.wrf-model.org/>
 - ¹³ Earth System Grid Federation home page, <http://esgf.org/>
 - ¹⁴ DCMIP homepage on CoG, <http://www.earthsystemcog.org/projects/dcmip/>
 - ¹⁵ IPCC Fourth Assessment Report, http://www.ipcc.ch/publications_and_data/publications_ipcc_fourth_assessment_report_synthesis_report.htm
 - ¹⁶ CMIP3 overview, http://cmip-pcmdi.llnl.gov/cmip3_overview.html
 - ¹⁷ Metafor homepage, <http://metaforclimate.eu/>
 - ¹⁸ Curator information, <http://www.earthsystemcog.org/projects/earthsystemcurator/>
 - ¹⁹ Lawrence, B. N., V. Balaji, P. Bentley, S. Callaghan, C. DeLuca, S. Denvil, G. Devine, M. Elkington, R.W. Ford, E. Guilyardi, M. Lautenschlager, M. Morgan, M.-P. Moine, S. Murphy, C. Pascoe,

H. Ramthun, P. Slavin, L. Steenman-Clark, F. Toussaint, A. Treshansky, and S. Valcke (2012). Describing Earth system simulations with the Metafor CIM, *Geosci. Model Dev.*, Volume 5, pp. 1–8.

²⁰ NCPP home page on CoG, <http://www.earthsystemcog.org/projects/ncpp/>

²¹ OpenDAP home page, <http://www.opendap.org/>

²² NetCDF homepage, <http://www.unidata.ucar.edu/software/netcdf/>

²³ ArcGIS homepage, <http://www.esri.com/software/arcgis>

²⁴ QGIS homepage, <http://www.qgis.org/>

²⁵ R homepage, <http://www.r-project.org/>