**Final Report for Period:** 08/2009 - 07/2010  **Submitted on:** 09/19/2010

**Principal Investigator:** Rugaber, Jon Spencer  .  **Award ID:** 0513635

**Organization:** GA Tech Res Corp - GIT

**Submitted By:**

Rugaber, Jon Spencer - Principal Investigator

**Title:**

Collaborative Research: Earth System Curator: Spanning the Gap Between Models and Datasets

## Project Participants

**Senior Personnel**

> **Name:** Rugaber, Jon Spencer
>
> **Worked for more than 160 Hours:**  Yes
>
> **Contribution to Project:**
>
> **Name:** Mark, Leo
>
> **Worked for more than 160 Hours:**  Yes
>
> **Contribution to Project:**

**Post-doc**

**Graduate Student**

> **Name:** Dunlap, Ralph
>
> **Worked for more than 160 Hours:**  Yes
>
> **Contribution to Project:**
>
> Rocky was the Graduate Research Assistant hired for this
> project. He is PhD student in the database area and is
> supporting the metadata definition portion of the Curator
> project.
>
> **Name:** Navarro, Angela
>
> **Worked for more than 160 Hours:**  Yes
>
> **Contribution to Project:**
>
> Angela was a research assistant who worked on the software
> engineering side of the project.
>
> **Name:** Markonda, Swetha
>
> **Worked for more than 160 Hours:**  Yes
>
> **Contribution to Project:**
>
> Graduate research assistant; programmer

**Undergraduate Student**

> **Name:** Ansari, Sameer
>
> **Worked for more than 160 Hours:**  No
>
> **Contribution to Project:**
>
> Sam was an undergraduate researcher who joined the project this summer and has conducted an experiment, in which the CCSM
> climate model was configured to run on
> a cloud computing instance.

**Technician, Programmer**

**Other Participant**

**Research Experience for Undergraduates**

<div align="center"><u>**Organizational Partners**</u></div>

**National Center For Atmospheric Research**

Cecelia DeLuca is co-PI on the Curator Project. She heads
the Earth Systems Modeling Framework group at NCAR.
Don Middleton from the Earth Systems Grid group is also
a co-PI.

**Princeton University**

The Geophysical Fluid Dynamics Laboratory at Princeton University
is a co-PI on the project.  V. Balaji heads the group there.

**Massachusetts Institute of Technology**

The Earth, Atmospheric and Planetary Sciences Department at
MIT is a co-PI on this project. Chris Hill and John Marshall are
our contacts there.

**University of Reading**

We are collaborating with Katherine Bouton and Lois Steenman-Clark of
the University of Reading. They are part of the Numerical Model
Metadata project. One of Curator's objectives is to produce a common
metadata representation for climate models.

**CERFACS**

CERFACS is a participant in the PRISM project (http://www.prism.enes.org/) a European effort that parallels
the Earth System Curator project. We are working with them
to define common metadata schema for modeling climate simulations.

**NOAA - National Oceanic and Atmospheric**

During the course of the project, our partners from NCAR moved to NOAA

<div align="center"><u>**Other Collaborators or Contacts**</u></div>

We have worked with the School of Atmospheric Sciences at Georgia Tech to define an
infrastructure for creating Climate Science tutorials for use by undergraduates.

We have corresponded with Robert Muetzelfeldt, Honorary Fellow, School of Informatics,
University of Edinburgh, Edinburgh, Scotland concerning the use of Prolog and other
logic-based languages for representing constraints in climate model metadata.

We are an active, but informal, partner in the METAFOR project
(http://www.metaforclimate.eu/), a newly funded European initiative, similar in scope and
purpose with our project. Both groups strongly intend to continue collaborating to
produce a set of common metadata schema for archival and retrieval of climate models
and datasets.

<div align="center">**Activities and Findings**</div>

**Research and Education Activities:**

1. Together with our Curator partners, we have developed XML schemas that describe climate science models. In particular, we have designed schema for models, components, interfaces, and grids. The schema were presented at the GO-ESSP annual meeting. (http://go-essp.gfdl.noaa.gov/)

2. Because the climate modeling domain is complex, multiple schemas are involved. Consequentially, we have examined the question of 'schema aggregation'. That is, what mechanisms are appropriate for combining schemas.

3. Schemas, by themselves, only provide a structural description of the underlying domain concepts. Hence, we have also provided an RDF-based glossary mechanism for presenting a more detailed description of the schema contents.

4. We have done some work exploring technologies (GWT, Spring and ZK) for use as part of the Curator user interface.

5. We have tested schema designs by using them to model partial climate models provided by our partners at NCAR and GFDL.

6. We have performed an initial analysis of the question of 'component compatibility'; that is, whether two repository components are capable of working together in a coupled climate model.

7. We have explored the use of program analysis tools, in particular, Understand Fortran. The goal is to extract information from existing climate model source code to populate schema instance documents.

8. We developed and refined curriculum materials for use in advanced software engineering classes. The materials use climate modeling as a motivation for presentation of software architecture concepts.

9. We have completed work on the technical infrastructure required to provide an on-line interactive tutoring system to be used to teach undergraduates and, eventually, the public about climate science. We have made this publicly available as a SourceForge project. (http://curatordb.cvs.sourceforge.net/curatordb/esc/tutorial/)

10. We have assisted in the development of the prototype Curator Data Portal (CDP), (http://curator.ucar.edu/home.htm) which provides access to climate data models and datasets.

11. CDP provides faceted access to its repository assets, and we have assisted in the definition of these facets and their expression in RDF.

12. We have surveyed mechanisms for automatically generating coupler components for linking independently developed climate simulations. In particular, we explored the use of the Bespoke Framework Generator (http://intranet.cs.man.ac.uk/cnc/projects/bfg.php) in support of this activity.

13. We have explored the relation of scientific computing with enterprise computing in order to understand the extent to which a web-services approach can be taken to proving access to Curator resources. In particular, we have conducted an experiment in which the CCSM climate model was configured and deployed in a cloud

computing environment.

14. In conjunction with NCAR, we have developed a meta-data upload scheme, taking XML output from model runs and transforming it for storage in both a SQL relational database and an RDF triples store.

15. With respect to automatic generation, we have developed a prototype generator as a plug-in to Eclipse. The generator has two features: 1) It allows a modeler to describe a pair of models as an EMF model; 2) it generates both a coupler and a driver for composing the models.

16. We have performed a domain analysis over the domain of coupling technologies and expressed it as a set of feature model diagrams.

**Findings:**
The current versions of our schema and architecture are available at the Curator web site (http://www.earthsystemcurator.org/).

**Training and Development:**
Mr. Dunlap has been working on the project since its inception. He is a fourth-year PhD student in the Database area. Over the course of the project, he has made numerous presentations to our collaborators and partners on the issues related to the schemas and tools we have developed. He has also worked with the undergraduate teams who have built a Climate tutorial system. During the course of the project, he passed his PhD qualifying exam and has entered candidacy.

Ms. Navarro has graduated with a Masters degree in Computer Science. While a member of the project, she used our newly developed schemas to model the ESMF demo application. She also investigated the use of program analysis tools to extract data from existing programs in order to populate instance documents.

Ms. Patnaikuni Markonda is a Masters student and new member to the project who has begun investigation the question of automatically generating coupler components for linking separately developed climate simulators.

Mr. Ansari has just joined the project. He has conducted an experiment in which he configured and deployed the CCSM climate model on the cloud.

**Outreach Activities:**
We attended the Global Organization for Earth System Science Portal (GO-ESSP) community meeting in June 2007 and presented an overview of our work to date.

We participate in weekly teleconferences with our American and European partners in an effort to obtain convergence in our schema development efforts.

We participated in the European Metafor project's M6 (month six) meeting, at which we made a presentation on Curator.

We presented a poster at the American Geophysical Union Fall meeting in San Fransisco.

We gave a talk on 'XML Schema and Constraints' at LogicBlox, Inc.

We have organized a workshop on 'Coupling Technologies for Earth System Modelling' to be held at CERFACS in Toulouse, France in December, 2010.

We have been invited to give a keynote presentation at the Community Surface Dynamic Modeling System Meeting to be held in San Antonio, Texas in October, 2010. The topic of the meeting is Modeling for Environmental Change. Our presentation will be on our feature modeling work.

## Journal Publications

Spencer Rugaber, "Design Studies in Software Engineering Courses", Proceedings of the Humboldt Science of Design Workshop, p. , vol. , (2007). Published,

Dunlap, R.;  Mark, L.; Rugaber, S.; Balaji, V.; Chastang, J.; Cinquini L.; DeLuca, C.; Middleton, D.; Murphy, S., "Earth System Curator: Metadata Infrastructure for Climate Modeling", Earth System Informatics, p. 131, vol. 1, (2008). Published,  10.1007/s12145-008-0016-1

## Books or Other One-time Publications

## Web/Internet Site

**URL(s):**
http://www.earthsystemcurator.org/
http://swiki.cc.gatech.edu:8080/curator
http://sourceforge.net/projects/curatordb

http://sourceforge.net/projects/curatordb/

**Description:**
The first URL is the main external facing site for the project.
The second is the Curator swiki for activity conversations and preliminary work products.
The third is our SourceForge repository for archiving our work products

## Other Specific Products

**Product Type:**

**XML Schema**

**Product Description:**

We have contributed to the development of four XML schema
+ Model - describes a coupled climate simulation program
+ Component - describes an individual simulation
+ CIAO - describes a component's interface
+ Gridspec - provides metadata for describing computational grids

**Sharing Information:**

These schema have been presented at various workshops as described elsewhere in this report. In addition, they are available at our web site.

**Product Type:**

**Software (or netware)**

**Product Description:**

Toolkit for building climate tutorials.

**Sharing Information:**

It is a SourceForge project.

# Contributions

**Contributions within Discipline:**

The Earth System Curator team has created a software environment for assembling, running, and archiving information about climate models. The idea is to make it easier for scientists to perform modeling experiments and to coordinate with each other on efforts such as Model Intercomparison Projects (MIPs) and Intergovernmental Panel on Climate Change (IPCC) assessments. The Curator project enables groups to prepare data for submission to the Program for Climate Model Diagnosis and Intercomparison (PCMDI) for IPCC analysis.

The first goal of the Curator project is to supplement and refine metadata schemas such as the Climate and Forecast (CF) and Numerical Model Metadata (NMM) conventions so that they better serve model analysis projects and so that they are complete and precise enough to serve as the foundation for software tools that can automate workflow tasks. The Curator portal is a distributed environment, based on these extended schemas, that enables modelers to archive and query models, experiments, model components, and model outputs; test the technical compatibility of model components; and assemble and run multi-component models. The Curator architecture includes:

1. A central CDP-Curator Portal that provides access to Earth system model component software, output datasets, and metadata describing these. The CDP-Curator portal will use the Community Data Portal (CDP) developed at NCAR.

2. Satellite sites that can assemble and run models, and archive experiments and data. The prototype for the Curator satellite site is the Flexible Modeling system Run-time Environment (FRE) and associated database at GFDL.

3. A lightweight interface that enables model components to be described and archived by groups that don't have a Curator software installation. Provider sites will use a schema based on Curator extensions to NMM.

The design for the Curator system architecture and other materials are available on the Earth System Curator website, http://www.earthsystemcurator.org.

**Contributions to Other Disciplines:**

**Contributions to Human Resource Development:**

Four Georgia Tech student, one in the PhD program, two in the Masters program and one undergraduate have worked on the project. Both Masters students has graduated, and the other two are still in their respective programs.

We have developed an on-line tutorial technology to help train undergraduates and the general public. This effort has so far included the participation of sixteen undergraduate students at Georgia Tech as part of the capstone design project.

**Contributions to Resources for Research and Education:**

Our schemas are intended to support the climate science community.
Our tutorial system is intended to support education about climate science.
We have developed an approach to Software Engineering class projects ('design studies')
  as described in one of the journal publications

**Contributions Beyond Science and Engineering:**

# Conference Proceedings

**<u>Categories for which nothing is reported:</u>**

Any Book

Contributions: To Any Other Disciplines

Contributions: To Any Beyond Science and Engineering

Any Conference