



Exploiting the Grid for collaborative ensemble studies using GENIE Earth System Models

A. R. Price (1), Z. Jiao (1), T. M. Lenton (2), G. Williams (3), P. J. Valdes (3), S. J. Cox (1), GENIE team (4)

(1) Southampton Regional e-Science Centre, University of Southampton, U.K.

(2) School of Environmental Sciences, University of East Anglia, U.K.

(3) School of Geographical Sciences, University of Bristol, U.K.

(4) <http://www.genie.ac.uk/people.htm>

(a.r.price@soton.ac.uk)

GENIE is a Grid-enabled framework for composing an extensive range of Earth System Models (ESMs) and simulating multi-millennial timescales, primarily for the study of ice age cycles and long-term human induced global change. Grid computing technology enables the flexible coupling of constituent models, execution of the resulting ESMs and the management of the data that they generate. We demonstrate how the Grid software deployed for the project is being used to execute large-scale simulations of an ESM that would not be feasible on a single computational resource.

Within the spectrum of GENIE models a simulation over multi-millennial timescales can require anything from 2 hours to ~1,500 hours of sequential compute time. To perform ensemble runs of the more computationally demanding models therefore requires a significant amount of resource. Whilst an individual project member may own sufficient resource on an exclusive basis, this may not be cost-effective if the resource is then only used in intermittent bursts with time in between idle as the experimental results are analysed, reported, and the next experiment designed. However it is very cost-effective if a multi-disciplinary, multi-site team such as GENIE, each of whom has similarly intermittent need for bursts of compute power, pool their resources.

The GENIE data management system provides programmatic access to a dedicated database and file repository from within the Matlab problem solving environment. Exploiting the database system we demonstrate how large scale simulations can be performed by staging intermediate results in the repository and using a number of

distributed “worker” clients to autonomously progress the model states. Through the common system interface and integration into the scientists’ familiar tools / environments we enable the project team as a whole to cost-effectively undertake much larger studies than could have been contemplated individually.