

# A National Scale Microsimulation of Disease Outbreaks

Douglas J. Roberts, M.S.<sup>†</sup>, Diglio A. Simoni, M.S.<sup>†</sup>, Stephen Eubank, Ph.D.<sup>‡</sup>

<sup>†</sup>*RTI International, Research Triangle Park, NC*

<sup>‡</sup>*Virginia Bioinformatics Institute, Blacksburg, VA*

## OBJECTIVE

Our objective in this research is to take advantage of a supercomputer grid (TeraGrid) to develop a distributed memory national scale agent-based model (ABM) to study disease outbreaks at the micro level. This has data needs at both the national data surveillance and the local community structure and outbreak levels.

## BACKGROUND

Distributed memory ABMs have been used in recent years to predict the spread of disease in human populations. In general, these ABMs have simplifying assumptions that enable the model to run on small clusters (for local-level models) or supercomputers (for national scale models). However, more detailed local-level discrete-event models (e.g. EpiSims) have demonstrated the added utility of detailed local ABMs to understand the dynamics and constraints of public health interventions during infectious disease outbreaks. This work extends these models to a national level, which is only possible given the extensive computer resources available through the NSF TeraGrid. However, such a model will have data needs that have implications for existing and future disease surveillance and health information systems.

## METHODS

The NSF TeraGrid is the world's largest and most comprehensive distributed cyberinfrastructure for open scientific research. It is a network of nine partner sites, integrated through high-performance network connections. Each facility is a high-end experimental supercomputer site. The TeraGrid includes more than 250 Teraflops of computing capability and more than 30 Petabytes of online and archival data storage. The challenge is to build a model that can be distributed across several sites and synchronized effectively between sites to produce a model that runs efficiently.

EpiSims models the movements of individuals (agents) and their contact patterns in a fine-grained simulation that is a realistic and detailed representation of the roads, buildings, and

the demographics of the community. Using the natural history of the disease, one can emulate the spread of infectious pathogens in this cybercommunity. To structure this model, various types of data are required, including: population demographics information, such as immuno-compromised state, age, ethnicity, income, etc.

## RESULTS

The use of the computational facilities in the TeraGrid presents a number of challenges due to the differing characteristics of the different sites: heterogeneity and temporal variation, scale and scope of the simulations, computing speed and memory requirements, and efficiency and sustained performance. In addition, dividing the discrete-event ABM model over multiple TeraGrid sites introduces computational challenges more complex than the distributed memory challenges within computer clusters. Important design elements that affect the grid-enabled ABM include: partitioning and distributing the problem, activity resolution, synchronization between local and global interactions, and global I/O. To accomplish this we use a grid-aware implementation of MPI called MPICH-G2 and IBM's General Parallel File System (GPFS) which is a high-performance shared-disk file system).

It is possible to model an area for a given disease outbreak and compare the model results with actual data from an outbreak in that same area, including attack rates, timing and magnitude of cohorts, etc.

## CONCLUSIONS

Complex grid-enabled versions of ABMs are possible, given the resources of the NSF TeraGrid. However, this requires new computational tools and access to community-level data collected on a periodic basis. Various articles describing community-specific EpiSims can be found at <http://www.midasmodels.org>.

Further information:

Doug Roberts, [droberts@rti.org](mailto:droberts@rti.org)

<http://www.midasmodels.org>