

# Joint Experimentation, Data Management and Analysis Enabled by Trans-Continentially Distributed Linux Clusters

Dan M. Davis and Robert F. Lucas  
[ddavis@isi.edu](mailto:ddavis@isi.edu) & [rflucas@isi.edu](mailto:rflucas@isi.edu)

Information Sciences Institute  
University of Southern California  
Marina del Rey, California

**ABSTRACT**  
**Of Oral Presentation**  
**For JFCOM Distributed Center for FMS**

There is a well-defined and critical need to produce quantifiable results, which are derived from simulations, to support transformation experimentation of the Joint Forces Command (USJFCOM). This is driving the creation of very large and geographically dispersed data collections. The Joint Experimentation Directorate (J9) of USJFCOM and the Joint Advanced Warfighting Project are conducting a series of Urban Resolve experiments to investigate concepts for applying future technologies to joint urban warfare. The recently concluded phase I of the experiment required, utilized and integrated multiple scalable parallel processors (SPP) sites distributed across the United States. These were hosted by the supercomputing centers at Maui (NHPCC) and at Wright-Patterson (ASC-MSRC) on a net including J9 at Suffolk, Virginia, Topographic Engineering Center, Fort Belvoir, Virginia and SPAWAR San Diego, California. This computational power had to be harnessed by scalable code in order to model the capability of futuristic sensor technology and the complexity of the urban environment. For phase I the simulation generated more than two terabytes of raw data at rate of >10GB per hour. The size and distributed nature of this type of data posed significant challenges in developing the corresponding data-intensive applications that manage and analyze them. Building on lessons learned in developing data management tools for earlier Urban Resolves, a next generation data management and analysis tool, called Simulation Data Grid (SDG), was developed and implemented. The design principles driving the architecture of SDG were

1. minimize network communication overhead (especially across SPPs) by storing data near the point of generation and only selectively propagating the data as needed, and
2. maximize the use of SPP computational resources and storage by distributing analyses across SPP sites to reduce, filter and aggregate.

The key implementation principle was to leverage existing open standards and infrastructure from Grid Computing. The developed system services interface with and were built on top of Open Grid Services Architecture standard and existing toolkits (Globus). SDG services include distributed data query/analysis, data cataloging, and data gathering/slicing/distribution. It is argued that SDG has proven to be a general-purpose tool useful for a range of simulation domains.

*This material is based on research sponsored by the Air Force Research Laboratory under agreement number FA8750-05-2-0204. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation thereon.*

*The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of the Air Force Research Laboratory or the U.S. Government.*