CEDPS: Center for Enabling Distributed Petascale Science

A DOE SciDAC Center for Enabling Technology

Semi-Annual Progress Report for the Period October 1, 2006, through March 31, 2007

Executive Summary:

This report summarizes ongoing collaborative work, overall progress, upcoming collaborations, and presentations carried out by CEDPS during the period October 1, 2006, through March 31, 2007. Significant progress has been made in each of the three CEDPS subareas: Data, Services, and Troubleshooting. Our primary collaborative focus has been on strengthening ties with the Open Science Grid (OSG) and Earth System Grid (ESG). In the next six months, we plan significant work to expand the number of end-users of CEDPS technology.

The Data area will provide key technologies for petascale science in DOE that will offer policy-driven, end-to-end placement of large-scale data as well as the services and workspaces required to process the data. One focus of the Data area is the Managed Object Placement Service (MOPS), which will provide reliable data movement with resource management. MOPS underpins the higher-level data placement services that will offer a policy-driven, end-to-end service to orchestrate the placement of large-scale data, a particularly acute problem for end users with large data sets, such as the HEP and climate application communities, the Spallation Neutron Source at ORNL, and the Advanced Photon Source at Argonne. Data placement services will improve the performance of DOE applications by reducing the time to stage data in and out of computational resources. These prototypes are on target for delivery by end of the year.

The Services area has extended work in two directions. The first is to provide a full set of distributed computing tools to the Python programming community to integrate scientific codes in the DOE community, including DANSE and climate codes. The second is to provide support for Grid computing protocols to allow access to additional DOE-wide resources, including OSG. Development has continued for pyGridware and CLWrapper, two of our Python tools that offer additional functionality to allow legacy applications to run in distributed environments. The HEP applications STAR and Alice are providing initial end-user requirements for building a proof-of-concept infrastructure to adapt virtualization technologies. As a result of those activities, we produced two releases of the Workspace service that allows applications to be deployed by using virtual machine technology, thereby avoiding costly deployment costs of the applications themselves.

The Troubleshooting area has completed a logging “best practices” document and the design of a centralized logging service based on syslog-ng. It is now working with the Open Science Grid on deployment of that service, as well as increasing the Trigger service functionality in support of Earth System Grid.

To learn more about the project, please visit our website at http://www.cedps.net.

1 Introduction

The Center for Enabling Distributed Petascale Science (CEDPS) will produce technical innovations to allow for (a) rapid and dependable data placement within a distributed high-performance environment and (b) convenient construction of scalable science services that provide for the reliable and high-performance processing of computation and data analysis requests from many remote clients. CEDPS will also address the important problem of troubleshooting these and other related ultra-high-performance distributed activities from the perspective of both performance and functionality. The project involves staff at Argonne National Laboratory, Fermi National Accelerator Laboratory, University of Wisconsin, University of Southern California, and Lawrence Berkeley Laboratory.

For the first six months of operation, the overall goals focused on putting in place the management structure for this five-site center; understanding current technology and interactions; and working with pre-existing collaborations to understand end-user requirements, as well as beginning to reach out to new communities. The three areas – Data, Services, and Troubleshooting – are carried out somewhat independently, with weekly management calls to review progress and coordinate work with applications groups as needed. Primary group discussions included design and prototyping, with some deployment to follow in the next six months.

In our first six months we held two all-hands meetings: a kickoff meeting in October 2006 (http://cedps.net/wiki/index.php/25-10-2006) and a midterm meeting in March 2007 (http://www.cedps.net/wiki/index.php/12-3-2007). We use a wiki to track outputs and coordinate work with these areas (http://www.cedps.net) and bugzilla to track milestones (http://bugzilla.mcs.anl.gov/cedps/).

2 Work with Applications and SciDAC Communities

In the first six months, we focused our application and outreach work on strengthening previous ties, especially with the SciDAC CET Earth System Grid (ESG) project and the SciDAC Sustaining and Extending the Open Science Grid: Science Innovation on a Petascale Nationwide Facility (OSG) and beginning to contact initial additional partners, with ORNL being the largest of these.

OSG, ESG, and CEDPS colleagues are working together to define areas of interaction and complementary work at a higher level in order to strengthen overall distributed science work within SciDAC. A lunch meeting at the Atlanta SciDAC PI meeting in February was held with Craig Tull and Walt Polansky, and a half-day meeting at SDSC in March was held with Thomas Ndousse.

To address the critical issues of ensuring reliable delivery of climate data to the climate community, we worked with ESG to identify opportunities for improved data delivery and problem detection in the ESG services. The Data area has presented CEDPS plans for the Managed Object Placement Service and higher-level placement services. The Troubleshooting area continues ongoing work with ESG in the extension and deployment of the Trigger service, which performs warnings on errors.

Our work with OSG includes contacts with both end-user applications and the infrastructure team. In terms of end-user applications, the Data area is working with several HEP applications that use Globus version 2 data tools to evaluate the use of current and future tools for data management.

The Services area has been working with the HEP STAR and Alice applications on developing workspaces and services allowing authorized clients to provision a platform ready to execute the application. The workspace-based approach is attractive to these communities because the complexity of installation and maintenance of their applications currently prevent the researchers from leveraging additional resources in an effective way: while at any given time resources may be available to run the STAR application, unless the application happens to be installed there, the resources cannot be leveraged in practice. Using workspaces allows us to deploy virtual machines preconfigured to run STAR on available resources. To verify this approach, we worked with Doug Olson at LBNL and the rPath company (specializing in configuration management) to produce images of STAR. We then dynamically deployed images of STAR and its dependencies using the workspace service installed on the TeraPort cluster at the University of Chicago, as demonstrated at SC06. We are continuing our interaction with the STAR experiment to define conditions that need to be met to bring the workspace-based provisioning model into production.

The Services area has also been working with the HEP Alice collaboration as well as with climate scientists to extend the workspace-based approach to their applications. The climate applications in particular require a very consistent environment for execution: minor variations in compiler or library versions can cause significant differences in the results produced by the applications. The workspace-based approach eliminates many of these problems because it provides a consistent configuration environment across platforms.

The Troubleshooting area has been coordinating with the OSG infrastructure group to deploy a centralized logging system first on the OSG Validation TestBed and then on the Integration TestBed, with plans for a full role out to follow.

The largest of our new collaborations is with ORNL, in terms of working with their infrastructure teams and with their end users. Jennifer Schopf and Dan Fraser spent two days visiting ORNL in February, presenting CEDPS work and meeting with applications groups. A continuation visit is planned (possibly in June), and work is ongoing to deploy GridFTP on the machines used by the SNS in order to simplify their secure data movement, as their experiments begin to come online and generate exceedingly large data sets. In addition, Keith Jackson and the Services area have been collaborating with DANSE (Distributed Analysis of Neutron Scattering Experiments) principals to adapt their distributed computing environment to pyGridWare, with a prototype expected in the next six months. Initial work will be targeted at enabling DANSE to take advantage of Open Science Grid resources.

Several smaller collaborations also have begun, in many cases through discussions that began at the SciDAC PI meeting in Atlanta. The Data area is exploring a collaboration with Garth Gibson's Petascale Data Storage Institute to see whether CEDPS tools can provide mechanisms for file systems to convey data using GridFTP servers for better performance, especially with how the data is mapped on disk, a feature of MOPS. The Data area has also begun discussions with Bob Lucas's Performance Engineering Research Institute to help identify possible application users with performance problems in the data space. The Services group has met with the fusion software development community at TechX, who are exploring the use of CEDPS tools to expose the new fusion code as a Grid service, including making their stand-alone Fortran code into a network accessible service using standard protocols. The Services group is also in discussion with the Community Climate System Model (CCSM) collaboration to meet their need of consistent environments in which to run their codes, since minor variation in compiler/library versions can cause significant differences in the results produced by the applications.

3 Milestones

We use bugzilla to track milestones and work progress. The server is located at http://bugzilla.mcs.anl.gov/cedps, and all current bugs can be viewed at http://snipurl.com/17l9p . The milestone numbers below refer to the bugzilla bug numbers.

3.1 Data Area

Based on our interactions with application communities and their requirements for end-to-end data management, the Data area will provide key technologies for petascale science in DOE that will offer policy-driven, end-to-end placement of large-scale data as well as the services and workspaces required to process the data.

One focus of the Data area is the Managed Object Placement Service (MOPS), whose goal is to ensure reliability and efficient resource management inside the object mover. Currently, failures can occur during transfers because of problems on the receiver side, such as disk overflows, or on the server side, for example if the server is overloaded with requests. MOPS will greatly reduce the probability of these types of errors, which in the era of petascale computing can save many days of wasted effort in a single use. These problems are particularly acute in application communities such as HEP and climate, as well as large experimental data producers such as the Spallation Neutron Source at ORNL, and the Advanced Photon Source at Argonne. We have begun to collaborate with the SciDAC Petascale Data Storage Institute to optimize data transfers using MOPS.

A second focus of the Data area is the design and implementation of higher-level data placement services, which will offer a policy driven end-to-end service to orchestrate the placement of large-scale data sets. Data placement services will determine where to store and replicate data based on the policies and requirements of application communities. The goal of the work on placement services is to make DOE application execution more efficient by more effectively staging data into and out of computational resources. MOPS technology underpins data placement services.

- ***MS1:*** *Work with application communities, including OSG, ESG, LIGO, and others, to understand their requirements and use these to drive service development.*

We have worked with all these application communities to understand their end-to-end data placement requirements. Interactions are ongoing. For example, with OSG, we have worked to understand the problems that have occurred with lack of resource management in GridFTP servers, and we will address this work with the MOPS server that provides resource management. We have also learned that OSG applications sometimes have trouble with staging data quickly off computational resources. This issue will drive some of our work on data placement. ESG and LIGO are both interested in the improved functionality that MOPS will offer over existing GridFTP, and both of these applications are discussing their long-term data placement requirements with us to help drive the design of our placement services.

- ***MS2:*** *Enable local resource management within GridFTP server: Prototype and implement basic capabilities within the GridFTP server for protecting the underlying system from resource exhaustion, in particular, to better manage network connections and memory usage.*

The GridFTP team at Argonne and the NeST teams at Wisconsin have created a baseline architecture and are currently developing the protocols and a prototype implementation that will allow basic memory and file system management in MOPS.

- ***MS3:*** *Prototype and implement NeST capabilities for managing the underlying storage component of GridFTP transfers. An important study will be to minimize the storage management impact on the GridFTP transfer rate.*

The Wisconsin team has made significant progress on integrating NeST into Condor as a precursor to integrating with GridFTP.

- ***MS4:*** *Combine enhanced GridFTP and NeST services into a MOPS 0.5 release with basic functionality. This release will be made available for users to start utilizing a basic MOPS functionality as soon as possible. We will work with OSG on evaluating, testing, and incorporating MOPS into the VDT.*

The GridFTP and NeST teams are on track to put together an initial working implementation of MOPS 0.5 by summer 2007. This version will include callouts made by GridFTP to ensure that file space is available when needed. Additional feedback mechanisms between NeST and GridFTP are being investigated to make sure the capability meets the anticipated user requirements for effectively reporting accurate utilization data, thereby allowing the system to effectively monitor resource utilization.

- ***MS5:*** *MOPS 1.0 release by October 2007 that builds on the MOPS 0.5 implementation and adds capabilities for exposing information on internally managed resources. We note that exposing resource information is only a first cut as some resource information may not be available via this mechanism. More sophisticated logic may be required to handle this problem at a higher level. MOPS 1.0 bundles will also be targeted for release in an upcoming OSG VDT release.*

The Argonne and Wisconsin teams are on track for a MOPS 1.0 release by October 2007.

- ***MS6:*** *Research optimization for transfer of Lots of Small Files (LOSF).*

The ANL GridFTP group has made significant progress on implementing optimizations for small files. A mechanism for streamlining multiple small files into an open channel all but eliminates the file transfer channel startup overhead and thereby enables bundles of small files (> ~ 50 KB) to transfer as fast as single large files.

Previously, performance could drop by an order of magnitude as soon as files less than several megabytes were encountered. These optimizations are targeted for the Globus Toolkit 4.1.2 development release, expected in June 2007, and for the Globus Toolkit 4.2 production release of GridFTP, expected summer 2007.

- ***MS7:*** *Investigate the integration of SRM and dCache with GridFTP service into the Managed Object Placement Service. Develop a plan for this work.*

Argonne and Fermi have met extensively to discuss the relationship of dCache and GridFTP and to identify a plan for this work, which includes GSI authentication caching. Fermi is also investigating the management of movement and storage resources, with the goal of controlling these resources and from this work defining the information that can be exported to higher-level services to enable them to make decisions on storage element selection for smart data placement.

- ***MS8:*** *Design and prototype implementation of a higher-level, policy-driven Data Placement Service that interfaces with the reliable distribution layer.*

The ISI group and the Wisconsin group have been working on the design of higher-level placement services of two types:

(1) A data placement service that focuses on moving data off computational nodes as the data sets are produced; this has been identified as a challenging problem for HPC applications that run on clusters or supercomputers;

(2) A data placement service that does asynchronous placement or distribution of data based on policies, for example, based on policies of a virtual organization that identify preferences for data locations or policies that maintain a certain number of replicas for availability or durability of data.

- ***MS9:*** *Design interface between higher-level data placement services and the reliable distribution layer and implement a prototype.*

The ISI group is on track to implement a prototype interface by October 2007.

3.2 Scalable Services

The Scalable Services effort consists of two major components. The frontend, developed at LBNL, will generate Web service interfaces turning existing applications into network services. The backend, developed at Argonne, provides services for flexibly provisioning resources to ensure on-demand service deployment. User requirements for this work have primarily come from HEP applications including ALICE, STAR, and CMS and ATLAS applications with their OSG edge service work.

- ***MS 10:*** *Develop an architecture document for dynamic resource provisioning using GRAM.*

This document is still in progress. We have had several design meetings, and have performed considerable prototyping work.

- ***MS 11****: Dynamic provisioning for HEP applications.*

We have developed a proof-of-concept infrastructure for dynamic provisioning using the Workspace service, assisted with the development of an image for the BNL STAR application, and used the resulting infrastructure to demonstrate dynamic provisioning of resources for applications. We subsequently applied the methods developed for the STAR application to use with AliEn, the ALICE collaboration's computation platform. So far, the infrastructure is deployed on one site, the University of Chicago TeraPort cluster. The requirements generated by those efforts are being folded into Workspace service releases. They include support for multiple VM partitions to make transfers more efficient, dynamic blankspace creation for application output data, and accommodation of the x86_64 architecture. We are also working on image improvement and hypervisor deployment to move these proof-of-concept efforts into production.

- ***MS 14****: Develop an architecture and roadmap for dynamic provisioning using virtual machine technology.*

We have developed an architecture and roadmap for dynamic provisioning using virtual machine technology based on user feedback on application requirements. Relevant designs include the following:

1. Configuration management tools (STAR, Alice, CCSM): better descriptions of images as well as better factoring are needed. Some information needed by services internal to the image is not available until deployment; we will leverage general configuration management tools.
2. Partition Management (STAR, Alice): VM images can be large (6-8 GB for STAR) but are composed of partitions some of which can be shared between multiple VMs. We need to develop methods for partition management and sharing that allow multiple VM deployments to not repeat transfer work or take up too much disk space.
3. Caching (STAR, Alice): internally transferring images or their partitions takes a long time; we are developing automated methods that would allow deployments to reuse as many images/partitions as possible by using caching strategies.
4. Virtual Clusters. Some applications rely on the presence of specific infrastructure in order to run: for example, STAR nodes need job submission infrastructure to enable users to submit jobs. We are developing methods allowing deployments to dynamically stand up such complete "virtual clusters" in addition to just application nodes.

- ***MS 15****: Library of Workspace service images.*

We have initiated the development of a Workspace Service Marketplace (http://workspace.globus.org/marketplace.html) which contains a library of Workspace images for scientific applications. The marketplace is modeled on similar efforts in industry by VMware and Amazon's EC2. It is intended to make it easier for scientists to find and download images needed to deploy their work on the Grid. The marketplace currently contains the images developed for application communities we actively engage with (STAR, Alice), along with other images useful for Grid computing, including the images of OSG edge services.

- ***MS 16:*** *Workspace service release.*

We have released two versions of the Workspace service, 1.2.1 in October 2006 and 1.2.2 in January 2007, each integrating features developed in response to proof-of-concept requirements as well as feedback generated from working with the STAR and Alice collaborations. Detailed change logs are available at the following websites: http://workspace.globus.org/vm/TP1.2.1/index.html and http://workspace.globus.org/vm/TP1.2.2/index.html.

- ***MS 17:** Network service requirements for pyGridware wrappers.*

We have begun to develop requirements for an automatically generated science service. We have worked with wrapping several molecular modeling codes and using those codes as part of a complex biological workflow. We have also discussed requirements with members of both the ESG project and the TechX-led fusion SciDAC2 project.

- ***MS18:** Develop pyGridWare-based portal generation services.*

PythonCLServiceTool was renamed to CLWrapper to emphasize the cross-language feature support. These features increase the applicability of CLWrapper to multiple target languages, as well as increasing its usability by adding portal generation capability, allowing existing Grid services to be easily adapted into a Web user interface. Work included the following:

1. Java client binding generation. A Java-based command line client thath behaves identically to the Python command line client was implemented on top of the client bindings. This enables the use of the client side of CLWrapper in pure-Java environments.
2. A JSR154-compliant servlet based on the Java client. This allows CLWrapper client functionality to be included in Java Web application hosting environments such as Tomcat.
3. A JSR168-compliant portlet based on the Java servlet. This allows CLWrapper client functionality to be embedded in portlet frameworks such as GridSphere.

This work in CLWrapper will inform further portal generation capabilities in pyGridWare based on TurboGears, a Python Web application framework.

- ***MS19, 20:** CLWrapper and pyGridWare instrumented with the NetLogger toolkit.*

Completed. CLWrapper and pyGridWare were instrumented according to the "best practices" document provided by the Troubleshooting area. This instrumentation logs all of the important calls in CLWrapper and pyGridWare including start and end of service operations, serialization and deserialization, and higher-level constructs such as service initialization, container operation, and notifications.

- ***MS21:** Adaptor classes for pyGlobus interfaces to legacy pre-Web Service tools.*

Initial work has been done on adding basic GridFTP support into pyGridWare to allow easy data transfers from a Web service environment. The basic support needed to add a pyGlobus style GRAM interface has been added, and work is ongoing to complete this interface.

- ***MS22:** pyGridWare support for additional high-level services.*

PyGridWare was enhanced to include a Delegation Factory Service and Delegation Service so that clients can use the pyGridWare container to delegate their credentials. Delegated credentials are required for the container to invoke services on the user's behalf. Client-side support for the Reliable File Transfer (RFT) service and the WSRF query service was also implemented. Initial work on supporting a GT4-style information service has begun, as has support for a GT4 compatible Reliable File Transfer Service. Support for these services will allow a pyGridWare container to be used interchangeably with a Java based container.

- ***MS24:*** *Customer/User Application Interfacing.*

We continued support of the LIGO project's use of pyGlobus by fixing bugs and implementing features at the LIGO community's request.

3.3 CEDPS Troubleshooting

The CEDPS troubleshooting team is developing tools to assist in the automated failure detection for Grid services, the end-to-end failure detection and analysis of user jobs, including log file discovery and management, and automated performance degradation detection for both Grid services and user jobs.

- ***MS25:*** *Log File Collection Service: This service will be used to gather middleware and application log files from multiple locations to a single host.*

We explored a number of open source tools and determined that syslog-ng (http://www.balabit.com/downloads/syslog-ng/) fills all our requirements. We tested it thoroughly and put together a set of recommendations for Grid deployment at (http://www.cedps.net/wiki/index.php/Syslog-ng). We are in discussions with OSG for initial deployments on the VTB.

- ***MS 26:*** *Write a Instrumentation "best practices guide": This is advice to middleware and application developers on what to log for maximal troubleshooting abilities. This includes what data should be included in the log files, and what format the log file should be.*

Completed. A document is available from http://www.cedps.net/wiki/index.php/LoggingBestPractices. This document was presented at the OSG and CEDPS all-hands meetings in March 2007.

- ***MS 27:*** *Develop a Event Log database. This database will be used to store log data and perform queries on log events to determine what was happening just before an error occurred.*

A prototype version is complete. It includes a tool to load log events into a SQLLite database. In the coming months we will replace SQLLite with PostgressSQL and test performance and scalability issues.

- ***MS28:*** *Instrument Globus and Condor based on the Logging Best Practices guide.*

Effort for this work within the Globus group has been earmarked for April 2007. Other middleware changes are still under discussion.

- ***MS29:*** *Develop additional MDS4 Trigger Service Information Providers.*

Work on incorporating ESG-specific information providers has been completed and included in the GT 4.1.0 release. Additional Information Providers, including one to track GRAM behavior for OSG, are under way.

- ***MS30:*** *Develop additional MDS4 Trigger Service action scripts (currently only send email).*

Current Trigger work is focusing on usability, ease of deployment, and easier setup of Triggers. Additional actions are planned as part of the Globus Toolkit 4.2 release, planned for summer 2007.

- ***MS31:*** *Deploy new services on OSG and ESG.*

Progress: We have been working extensively with OSG, please see previously mentioned agreements. We have also had discussions with ESG.

- ***MS32:*** *Outreach to Grid application developers to instrument their applications.*

We have begun discussing this with the STAR project and with the SRM/DRM developers.

- ***MS34:*** *Log file anomaly detection service.*

A prototype is completed and was demonstrated at the CEDPS midterm meeting in March 2007. This is part of the planned deployment with OSG.

4 Future Work with Applications and SciDAC Projects

We will continue our outreach efforts to additional SciDAC projects and application groups in the coming year. We have requested time to present a tutorial of CEDPS tools at the SciDAC PI meeting in June. As part of our ongoing collaboration with ORNL, we plan to set up another two-day visit, which would include a half-day tutorial for Globus services and a half-day presentation of CEDPS tools and plans; participants would comprise CEDPS area leads and ORNL project personnel, including John Cob and Steven Miller from the APS.

In addition to existing collaborations discussed in Section 2, the Data area plans to explore a possible collaboration with the combustion application group based at LBNL. In addition, the Data area scientists will work with current users of the Globus Data Replication Service to see whether these users have additional requirements for higher-level data placement services.

The Services area will meet with ESG to explore exposing several of their major climate codes as Grid services using pyGridWare. Since the climate community is moving more and more to Python as their standard scripting language, this strategy will allow them access to all of their major codes from their preferred environment. In addition, Workspace service work will be extended for ALICE and CCSM.

The Troubleshooting area plans to start working with STAR and other OSG users, as well as Shoshani's Scientific Data Management Center for Enabling Technologies in the next year.

5 Presentations

Ann Chervenak, “Next Generation Data Services,” Fermi National Laboratory, October 27, 2006

Tim Freeman and Kate Keahey, “Virtual Workspace Appliances,” SC06 Booth Presentation, November 2006
http://workspace.globus.org/papers/workspace_appliances_sc06_booth.pdf

Kate Keahey, “On-Demand Virtual Workspaces: Quality of Life in the Grid,” 5th Meeting of Spanish Initiative in Grid Middleware, November 2006
http://workspace.globus.org/papers/on_demand_workspaces_granada.ppt

Ann Chervenak and Dan Fraser, “Data Services: Future Directions,” TeraGrid Data Meeting in San Diego, January 9-11, 2007.

Dan Fraser and Ann Chervenak, “Data Services: Future Directions,” Earth System Grid meeting in Boulder, CO, January 17-19, 2007.

Jennifer M. Schopf, “CEDPS and CDIGS: Two Globus Projects,” Middleware And Grid Infrastructure Coordination (MAGIC), National Science Foundation, Arlington, VA, February 7, 2007. http://www-unix.mcs.anl.gov/~schopf/Talks/cedps-nsf-feb07.ppt

Jennifer M. Schopf, “SciDAC Center for Enabling Distributed Petascale Science,” Oak ridge National Laboratory, Oak Ridge, TN, February 8, 2007. http://www-unix.mcs.anl.gov/~schopf/Talks/cedps-ornl-feb07.ppt

Ann Chervenak, “Data Services in the SciDAC CEDPS Project,” OSG Consortium All Hands Meeting, March 6, 2007.
https://indico.fnal.gov/contributionDisplay.py?contribId=90&sessionId=53&confId=468

Brian Tierney, “Logging Recommendations for Effective Troubleshooting,” OSG Consortium All Hands Meeting, March 6, 2007.
https://indico.fnal.gov/contributionDisplay.py?contribId=120&sessionId=13&confId=468

Ian Foster, “System-Level Science: Scientific Exploration & IT Implications,” Keynote, Workshop on Wireless Networking, Automated Information Processing, and Web & Grid Services, Puerto Rico, February 4, 2007.

Ian Foster, “Grid,” IFIP Summer School on Software Engineering and Computer Science, Gordon’s Bay, South Africa, March 26, 2007.

Ian Foster, “Service-Oriented Science: Scaling eScience Impact,” Distinguished Lecture, Louisiana State University, November 27, 2006. Also presented as Keynote, Web Intelligence Conference, Hong Kong, December 18, 2006; and at the University of Capetown, March 29, 2007.