The Particle Physics Data Grid (PPDG): From Fabric to Physics, Final Report, July 2006

Table of Contents

A Introduction

The Particle Physics Data Grid (PPDG) Collaboration was formed in 1999 by particle physics and computer science groups from several laboratories and universities. DOE Next Generation Internet (NGI) funding for PPDG sponsored a program of work to test that "an infrastructure built upon emerging network and middleware technologies can meet the functional and performance requirements of wide area particle physics data analysis". NGI funding was followed by MICS base-program funding and then by five years of SciDAC funding from the MICS, HEP and NP program offices. PPDG progressed from successful testing of the "can meet" hypothesis to the current state where PPDG-hardened middleware *does meet* the requirements of a wide range of particle physics applications.

The Particle Physics Data Grid (PPDG) Collaboration has deployed, used and extended production Grid systems — vertically integrating experiment applications, Grid technologies, Grid and facility computation and storage resources to provide effective end-to-end capabilities. In 2005 PPDG joined with the NSF-funded iVDGL, US LHC Software and Computing projects, DOE Laboratory facility and other groups to build, operate and extend their systems and applications on the production Open Science Grid.

PPDG is a collaboration of computer scientists with a strong record in Grid technology, and physicists with leading roles in the software and network infrastructures for major high-energy and nuclear experiments. The goals and plans, guided by the immediate and medium-term needs of the physics experiments and by the research and development agenda of the computer science groups, resulted in a dramatic transition in the manner and capabilities of performing scientific computing for many high-energy and nuclear physics experiments. The transition from a centralized facility based approach to data intensive computational workloads distributed across shared facilities has occurred in large part due to the accomplishments of the PPDG project.

In the first years of its SciDAC program, PPDG recognized that success required progress on a broad front, and thus the harnessing of enthusiasm of many more physicists and computer scientists than those supported by PPDG funding. The "flower plot" developed by the project early on illustrates the relationship between PPDG and the wider physics and computer science communities that have been involved. The major legacy of PPDG is this transformative change in the close collaboration between people from different science disciplines, physics and computer science groups, and projects, iVDGL, GriPhyN and PPDG, towards the vision of globalization of scientific computational analyses based on a common cyberinfrastructure.

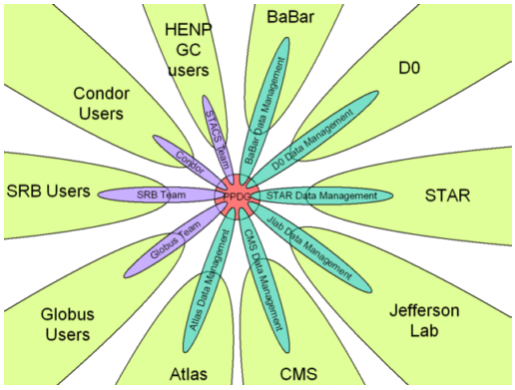

Figure 1: PPDG organizational locus

To widen and motivate participation, PPDG initially steered the formation of "project activities", each involving computer science groups working with physicists who were enthusiastic to derive benefit from beginning to deploy Grid technologies. The project activities resulted in Grid tools being used to facilitate science and at the same time undergoing rapid hardening and appropriate redesign spurred by the information exchanged between physicists and computer scientists. Project activities were complemented by "cross-cut activities", for example the collaboration with DOE Science Grid to set up a certificate/registration authority for PPDG, and the PPDG SiteAAA (Authentication, Authorization and Accounting) project that attracted some incremental funding.

B Common Services Accomplishments

As PPDG approached the end of its first three years as a SciDAC project, it became possible to focus its community towards the adoption of a standardized tool subset implemented on a pooled, sharable Grid fabric. PPDG oriented its future plans around the implementation, with its partners, of the US Open Science Grid. To support this evolution, the management effort and the resources devoted to crosscut activities were greatly increased by creating the "PPDG Common Project". The common project has focused on the integration and testing of the OSG Software Toolkit, including the development of essential additional functions such as security policy and architecture, identity management and accounting. Many of the accomplishments of PPDG are reflected in the use and continued plans for evolution of the common software, services and systems that have been developed.

B.1 The Common Project

The PPDG Common Project, which includes people from 8 different organizations, has provided the core of the technical work towards common and generally usable Grid technologies for the stakeholders. The work of the Common Project has included: extension and deployment of Grid security and authorization components; testing and deployment of storage management implementations with common interface specifications and semantics; monitoring and

information services; grid accounting; testing and validation tools and component and system integration.

Figure 2: Common Project Coordinators 2006

All of the PPDG Common Project activities include collaborators from outside PPDG and have had to compromise between the needs of individual VOs and the greater grid community. The PPDG Common Projects have contributed directly to the Open Science Grid program of work, and many deliverables are included in the OSG software stack:

- The Storage Resource Manager (SRM) Tester has been integrated into the OSG operations toolkit. It currently tests conformance to SRM specification v1.1 and v2.1 and is deployed through the Virtual Data Toolkit (VDT).
- Authorization Services: The Authorization service PRIMA has been deployed. Recent extensions have included the port to 64 bit Linux platforms and integration into the GT4 Web Services GRAM framework. Identity mapping service GUMS has been deployed at all US LHC and STAR sites. Sites configure GUMS to implement configurable local identity mapping. Storage authorization callout gPLAZMA is being included in the next release of the Storage Element dCache implementation.
- The Grid validation and tester "Grid Exerciser" has been developed and is run on the OSG Integration Testbed, CMS datagrid, and for other Virtual Organizations (VOs) such as Nanohub. The lessons learned have been included into the Condor Grid Universe (Condor-G) client.
- A Resource Selection Service based on matchmaking with Condor ClassAds has been developed for DZero and is being stress tested and deployed on the OSG Integration Testbed (ITB).
- An Edge Services Framework based on the Xen Virtual Machine technology is being developed for US ATLAS and US CMS deployment.
- A Clarens based Service Discovery Service has been developed and deployed in test on the OSG. It is currently used by the US CMS as a catalog of the installed application software.
- A Grid Accounting infrastructure, Gratia, has been designed, implemented and deployed in test and will be released for OSG in the next few months.

B.2 Open Science Grid

The creation of the Open Science Grid (OSG) Consortium has resulted from the collaboration started between the PPDG, iVDGL and GriPhyN projects and stakeholders, and extended to include other contributing groups. The OSG is a consortium of computing facilities in DOE and NSF, computer scientists, information technology engineers, physicists, biologists, astrophysicists and researchers from other domains to maintain and operate a premier distributed facility, to provide education and training opportunities in its use, and to expand its reach and capacity to meet the needs of the stakeholder and other scientific organizations.

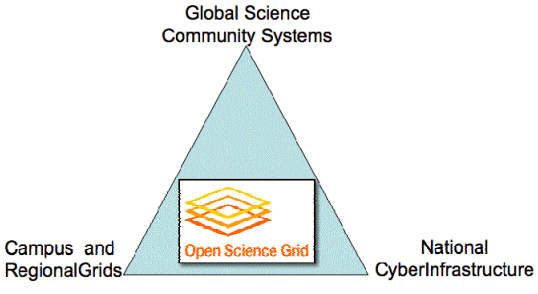

Figure 3 OSG Ecosystem

OSG provides a sustained, common, shared distributed infrastructure to access a large number of compute and storage resources through both production and research networks. The application research groups work closely with the distributed facility organizations towards effective use of the end-to-end system and to ensure the stakeholder needs are met in a timely and robust manner. OSG provides the US facility in the global cyberinfrastructure of most of the research groups.

OSG is also active in the areas of Campus and Regional Grids and in working in the area of interoperability between Grid infrastructures —in particular with the TeraGrid and Enabling Grids for EsciencE (EGEE).

In the final 2 years of PPDG much of the projects focus has been towards building the OSG and providing operational services and capability extensions to meet the stakeholders' needs.

Figure 4: OSG Consortium Meeting 2006

B.3 The Virtual Data Toolkit

The Virtual Data Toolkit (VDT), initially developed by GriPhyN and supported by iVDGL, has become the software packaging and distribution mechanism for PPDG sponsored deliverables, common middleware in use by the PPDG and Open Science Grid stakeholders. The added value of the VDT is the integration, testing and configuration done as part of the packaging and support. A lot of effort is required to provide an easily installed, configurable and integrated software stack of many diverse components from independent software developers. The value of a central group doing this work for all cannot be overestimated. The VDT depends on the NSF National Middleware Initiative (NMI) for the core middleware components —including Condor and Globus — and the comprehensive NMI build and test infrastructure.

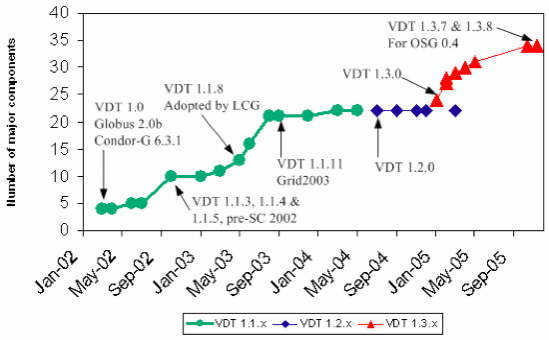

Figure 5: VDT downloads during its first 3 years

VDT provides procedures for requesting new components and providing a framework for the support and evolution of the software by the development and operations groups. VDT was early on adopted as the underlying middleware for the European physics-focused grid projects including the Enabling Grids for EsciencE (EGEE) and Worldwide LHC Computing Grid. Work is ongoing with TeraGrid to ensure a consistent and common base of middleware with the Open Science Grid.

B.4 Common Certificate Authorities and Management

The PPDG Registration Authority provides a common interface for the request and management of PKI Certificates from the DOE Grids Certificate Authority operated by ESnet.

This X509 based identity management infrastructure continues to be at the core of the PPDG and OSG security infrastructure enabling users to access and utilize the distributed computing

resources. The PPDG, FNAL and iVDGL RAs today issued most of the certificates to participants in Open Science Grid, amounting to about 80% of the total DOEGrids certificates. Additionally, PPDG effort has been used to establish the OSG RA with a scope to cover any participant in OSG. PPDG additions for usability and deployment include MyProxy as a certificate repository as well as utilities for the management of host certificates and certificate revocation services, which have been included in the VDT.

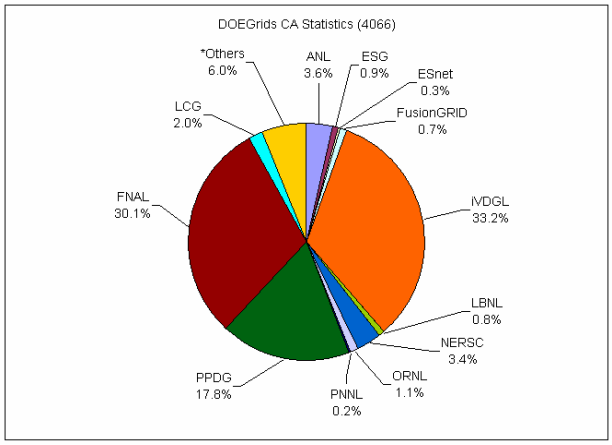

Figure 6: Active certificates issued, by RA

B.5 Interoperability and Partnerships

A PPDG deliverable has been collaboration across project and national boundaries that reflects the global nature of the stakeholder science collaborations. PPDG was a major contributor to early efforts of interoperation and has sustained these commitments through to the Open Science Grid. These include the development of common interface to storage services (Storage Resource Management specification), information schema (Glue-Schema), and interoperation activities in WorldGrid, WLCG, and the experiment data and job management systems. This work continues together with EGEE, TeraGrid, Pragma, Naregi and NorduGrid in the Grid Interoperability Now (GIN) program of work. Interoperability with TeraGrid got a boost, which the PPDG Common Project coordinator Dane Skow recently became Deputy Director of the TeraGrid Grid Infrastructure Group (GIG).

Other ongoing collaborations include those with the Internet End-to-end Performance Monitoring project which develops and deploys comprehensive monitoring of network connectivity and end-to-end performance for sites involved in High Energy Nuclear and Particle Physics, and with the MonaLisa service and resource monitoring project.

C Science Accomplishments

The science accomplishments and results from PPDG include simulation and experimental results for the physics collaborators and computer science deliverables for the middleware technology groups. As stated above, PPDGs core mission is to enable production quality end-to-end data analysis systems for its stakeholders based on common software and services. Thus much PPDG effort was focused on ensuring the integration and adoption of grid technologies into each of the much (much) larger experiment software and computing organizations. Below are some representative samples of the accomplishments from each of the participating organizations that are part of PPDG.

C.1 ATLAS

PPDG—together with iVDGL and GriPhyN and now OSG —has enabled US ATLAS to participate fully in the simulation distributed production activities and developments for experimental event data analysis. The MAGDA data management

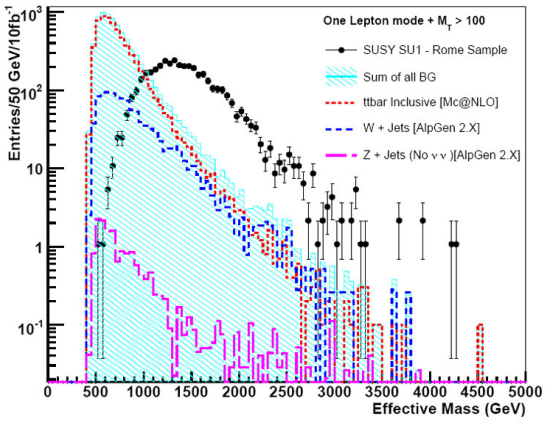

Figure 7: ATLAS Susy Simulation Results from Grid Computing

system was used for the first three years of data simulation production; the current PANDA system will also be used for the actual data processing and analysis when beam comes in 2007.

As a result of the availability of a common distributed infrastructure in the US through which ATLAS jobs can be opportunistically executed on shared resources to date 20% of the experiment's data simulation has been done in the US and new results in understanding underlying physics potential and measurements at the LHC have been achieved.

One such is the study the background processes supersymmetry (SUSY). These studies involve generation and simulation of complex Standard Model (SM) processes. In the lepton channel example shown, expected to provide a clean discovery channel in the early months of the LHC's running period, there is a clear enhancement of the SUSY signal over all SM background processes.

C.2 BaBar

The BaBar experiment at SLAC routinely moves all the data collected from the accelerator to computing centers in France, Italy and England. BaBar and the Storage Resource Broker (SRB) groups have worked throughout the lifetime of PPDG to deploy, harden and extend SRB and the BaBar application middleware to provide the distributed catalogs and data management services needed to sustain and increase the data analysis and distributed systems. Positive and important results of this collaboration are the development of distributed meta data cataloging system (MCAT) in SRB, the extensions of the meta-data in the SRB catalogs, and the ongoing reliance of BaBar physicists on the distributed datasets throughout Europe and the US.

C.3 CDF

CDF was a relatively latecomer to the world of PPDG. The experiment has focused on providing remote analysis facilities (DCAFs) and then migrating to Condor based Grid based computing for data simulation and analysis. CDF depends on PPDG supported extensions to the Condor middleware and services that support a fully Kerberos based security system, "Glide-In" job scheduling at the remote resources and Grid Connection Broker (GCB), and is testing, Computing on Demand (COD).

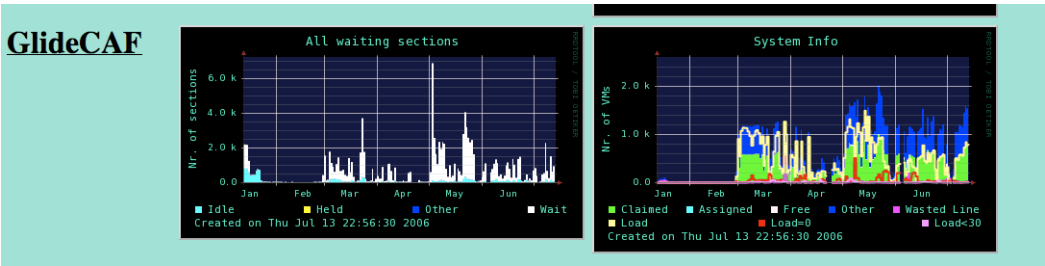

Figure 8: Jobs running on GlideCAF over past year

C.4 CMS

Using PPDG as well as many other contributions, CMS has developed and uses a Grid based distributed system for simulation data production and analysis. Currently the experiment's CRAB distributed analysis tool provides all CMS users with access to all experimental and simulated data samples. Utilization information from sites worldwide is used to decide which jobs are sent to which site uniformly across the EGEE and OSG infrastructures. This brokering relies on the

interoperation of the EGEE and OSG information providers, recently achieved by a collaboration between the two Grid projects.

To simulate user analysis, Job Robots that use CRAB in an automated way have been able to reach the goal of 12,500 analysis jobs per day on EGEE and OSG resources. As the user analysis tool is used directly by the robots, the results are directly applicable to the real user analysis workflow. In just a few days, thousands of jobs have been completed with efficiencies around 90%.

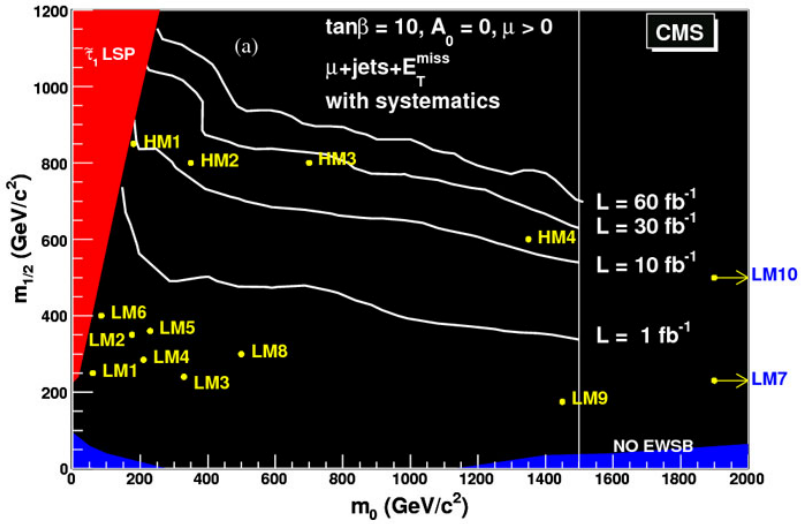

Figure 9: Part of a 2D scan of the CMS discovery potential for supersymmetry.

Example analyses are those that have been done to investigate the potential of the CMS Detector to discover supersymmetry and to find selection criteria in the generated datasets that isolate the supersymmetric signal from the Standard Model backgrounds, analyzed them. The figure above shows their results as a two-dimensional scan of different supersymmetry mass parameters, where each contour represents the discovery reach for a different amount of collected data.

C.5 D0

D0 adopted a globally distributed computing model from the beginning of its preparations for data taking at the Fermilab Tevatron, developing first versions of the Sequential Analysis with MetaData (SAM) system as PPDG was being proposed. Initial versions of SAM used experiment-developed middleware and during the lifetime of PPDG the experiment has worked with the computer science groups to extend and adapt the software on both sides, and move SAM steadily towards the use of common services.

All D0 physics results now rely on simulation and data processing on the distributed system, which now consists of up to 25 sites as far away as India and China. Exciting new results such as the measurement of B meson mixing are being published by the experiment.

Specific PPDG accomplishments have been the development of the Job and Information Management System (JIM) which relies on extensions to the Condor components for resource selection and a three-tier "client-server" job dispatching architecture, hardening of Globus GridFTP and GRAM for the reliable transfer of

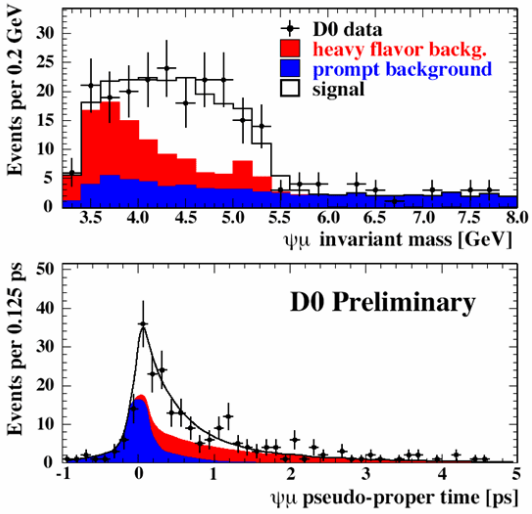

Figure 10 D0 preliminary results on B Mesons

100s of TBs of data and execution of order 100,000s of jobs, and extensions to SAM to take advantage of these.

C.6 Jefferson Lab

Much of Jefferson Lab's work on PPDG has focused on the design and implementation of the Storage Resource Manager (SRM) service and on the boundary between the SRM and the facility. An Application Program Interface (API) was developed for Jasmine, the Jefferson Lab Mass Storage System, to allow for the development of multiple Storage Resource Manager (SRM) services that are loosely coupled with, instead of a part of, Jasmine. Jefferson Lab developed an SRM based on version 2 of the functional specification for use by the laboratories LQCD collaboration and a functional SRM version 3 prototype that provides a Web Services interface to external clients.

In late 2005 Carnegie Mellon University (CMU) used the Jefferson Lab SRM version 3 prototype to transfer 11TB of CLAS raw data to CMU for data reduction and analysis. As a result, they were able to reduce the data down to 600GB and their processing time by about a factor of 3.

C.7 STAR

Through PPDG contributions STAR has been sustaining production data movement services between Brookhaven National Laboratory in New York and Lawrence Berkeley Laboratory in California for several years allowing next day data availability at remote sites. This was enabled through the experiment's collaboration with the Storage Resource Management group and transfer of more than 5 terabytes a week between the sites is ongoing.

Additional accomplishments are the development of a software package to provide a constant interface to the ever-evolving dynamic hardware and software that defines grid computing. The STAR Unified Meta Scheduler (SUMS) provides a simple and elegant definition of a physics user analysis and translates that into the required commands to allocate disk storage, locate data sets and break tasks into many processes that can run in mass parallel, launch jobs on the grid, and return the results to the user.

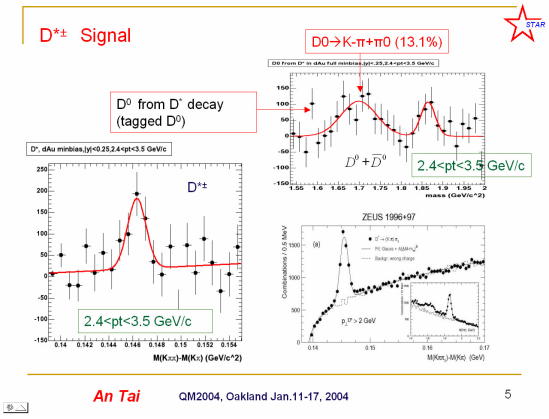

Figure 11: Early results in Charm production from STAR in 2004

A collaboration between STAR and the Scientific Data Management (SDM) Center developed a tool (the Grid Collector) for efficient selection across a billion indexed objects which greatly speeds up the analysis of STAR, opening the avenue for the search for rare events in large data samples. The shortened turnaround time for analysis of the STAR Nuclear Physics Experiment data produced early results in the first direct measurement of open charm production at RHIC

C.8 Condor

Many extensions and developments have taken accomplished in the Condor project components as a result of the deployment of the "real-life" systems of the PPDG physics collaborations. These developments have occurred in the core Condor job scheduling system especially in the

security and policy services, ClassAds, the workflow DAGMAN components, and in the development of new services such as "Computing on Demand", the Grid Exerciser (grid site testing application), the Grid Connection Broker, Condor-G 3-tier architecture, contributions to the European EGEE project, and more.

C.9 Globus

PPDGs early use of the Globus GIS security infrastructure, the GridFTP file transfer services, and the GRAM job execution middleware contributed significantly to their hardening and evolution to meet production needs of scientific applications. Later use of MDS-2, RLS and other components contributed to their evolution and production deployment. The PPDG experiment requirements were important inputs to the continued evolution of CAS, RFT as well as other Globus components.

C.10 Storage Resource Broker

As described above the SRB group has worked closely with the BaBar experiment to extend and harden the software and provide commonly reusable components for the general scientific community. . The SRB technology received an Internet2 Driving Exemplary Applications IDEA award for its use in the NARA transcontinental persistent archives prototype on April 26, 2006 in Washington, DC.

C.11 Storage Resource Manager

PPDG has had a very close working relationship with the Storage Resource Manager (SRM) collaboration. SRM V1, V2 and now V3 specifications have been written and adopted by US and European Grid accessible storage implementations, and SRM is a baseline service for the WLCG Collaboration. PPDG supported implementations of SRM are the DRM at LBNL, JASMINE at JLAB and SRM-dCache at Fermilab. A common SRM client is included in the Virtual Data Toolkit as well as the SRM-DRM server implementation. The LHC Service Challenge data distribution of 200MB/sec to the ATLAS and CMS US Tier-1 facilities is handled through the SRM interfaces.

Additionally SRM has developed the SRMTester, an independent test-suite for any SRM server implementations based on SRM interface specifications. The SRM-Tester has been deployed through the Virtual Data Toolkit (VDT), and it has been used in the Global Grid Forum (GGF) Grid Interoperability Now (GIN) SRM interoperability tests.

C.12 Grid Analysis Environment

Several of the Caltech Grid Analysis Environment (GAE) components have been developed through PPDG. GAE components are based on the Clarens toolkit and the MonALISA monitoring framework, funded by USCMS, and include the monitoring and accounting system used on OSG, application services using Clarens, specifically a service discovery service and job monitoring service. These are part of the Virtual Data Toolkit and available for use on the Open Science Grid. Clarens in also in use by the National Virtual Observatory, HotGrid and the LambdaStation bandwidth reservation system and is part of the CMS software distribution.

D PPDG Issues

The early estimates of "large-scale" development efforts that would be needed to deliver production quality Grid systems has proven largely correct. International efforts on Grid

development and deployment were funded below the "Microsoft-"scale, and were rendered more "interesting" by the absence of a single top-down management, let alone co-location of people. PPDG, GriPhyN, iVDGL, DataGrid, EGEE etc. addressed these issues by devoting significant resources to inter-project communication and coordination, resulting in international successes like the "GLUE schema". Overall, entropy has been well contained without stifling inventiveness. However, an inevitable consequence of the funding levels has been that a broad user community has been an essential part of the product hardening team. This approach has worked well in the particle physics community where research teams comprise hundreds or thousands of scientists, but it is not automatically valid for sciences where research teams are much smaller.

E Continuing Impact from the Project

PPDG lives on in the Open Science Grid project and Consortium. PPDG PIs Miron Livny and Ruth Pordes, coordinator Doug Olson, and PPDG team leads from Globus, SRM and all of the PPDG experiments are members of the OSG management teams who will continue their work towards a universal usable cyberinfrastructure for a broad range of scientific endeavors, and towards the education and training of the next generation workforce in research and distributed computing techniques. All the experiments active in PPDG are stakeholders in the OSG and continue to use and request extensions to the Grid technologies provided by the PPDG computer science groups.

OSG is now the core fabric for the future of US high-energy physics, and is already the principal resource for the majority of US experiments. OSG is already showing how other sciences can benefit from Grid technology and from joining a consortium that operates a sharable fabric. In OSG, the presence of well-organized communities of users and Grid sites, such as the high energy physicists and their computing facilities, has helped to bring services up to production quality for all science users.

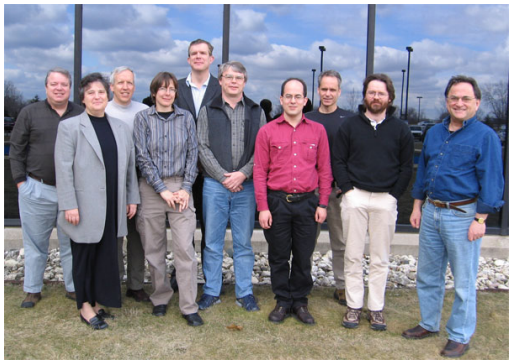

Figure 12 Members of the OSG Executive Board

PPDG has created collaborations between physicists and computer scientists that are themselves of lasting value. The PPDG computer science teams assert that there is exciting computer science in the collaborative research advancing from "proof-of-concept prototype" to middleware that meets the scientists' needs for function and robustness.

Finally, from PPDG's original strategy of "project activities" many valuable applications arose that continue to benefit science without being part of the core toolset of the OSG. For example, intercontinental replication of databases using SRB, pioneered as a PPDG activity, continues to serve the BaBar experiment and is proving well suited to the needs of biologists and other scientists. Many of the PPDG project activities now have a life of their own and will continue to bring benefit to science

F Published Papers based on PPDG Work, by date.

The participating scientific groups have all published papers that relied on deliverables from PPDG. As a middle-tier computing infrastructure project invariably the specific references to PPDG benefits and work are not called out.

Mehnaz Hafeez, Asad Samar, Heinz Stockinger, *"A DataGrid Prototype for Distributed Data Production in CMS"*, VII International Workshop on Advanced Computing and Analysis Techniques in Physics Research (ACAT2000), October 2000.

"Data Management in an International DataGrid", H. Newman, et al., IEEE, ACM International Workshop on Grid Computing [Grid'2000], 17-20 Dec. 2000, Bangalore, India.

Asad Samar, Heinz Stockinger, *"Grid Data Management Pilot (GDMP): A Tool for Wide Area Replication"*, IASTED International Conference on Applied Informatics (AI2001), Innsbruck, Austria, February 2001.

Proceedings of Computing in High Energy and Nuclear Physics Conference 2001 (CHEP'01), Beijing, China, Sept. 2001

"Grid Technologies & Applications: Architecture & Achievements", Ian Foster

" Jefferson Lab Mass Storage and File Replication Services", Ian Bird, Ying Chen, Bryan Hess, Andy Kowalski, Chip Watson

"Globus Toolkit Support for Distributed Data-Intensive Science", W. Allcock, A. Chervenak, I. Foster, L. Pearlman, V. Welch, M. Wilde

"Data Grid Services in STAR, Initial Deployment: Site-to-Site File Replication", D. Olson, E. Hjort, J. Lauret, M. Messer, J. Yang

"PKI and Alternative Security Architectures for Grid Computing", Robert Cowles

"SAM and the Particle Physics Data Grid", Lauri Loebel-Carpenter, Lee Lueking, Carmenita Moore, Ruth Pordes, Julie Trumbo, Sinisa Veseli, Igor Terekhov, Matthew Vranicar, Stephen White, Victoria White

"Resource Management in SAM- The D0 Data Grid", Lauri Loebel-Carpenter, Lee Lueking, Wyatt Merritt, Carmenita Moore, Ruth Pordes, Igor Terekhov, Julie Trumbo, Sinisa Veseli, Matthew Vranicar, Stephen P. White, Victoria White

"CMS Requirements for the Grid", K. Holtman, et al.

"Storage Resource Managers: Middleware Components for Grid Storage", Arie Shoshani, Alex Sim, Junmin Gu, MSS, 2002

"Interfacing interactive data analysis tools with the GRID: the PPDG CS-11 activity", D. L. Olson and J. Perl, Proceedings Of The VIII International Workshop On Advanced Computing And Analysis Techniques In Physics Research, Moscow, Russia, 24 - 28 June 2002, NIM A502 (420-422), April 2003

"The SAM-GRID project: architecture and plan", A. Baranovski, G. Garzoglio, H. Koutaniemi, L. Lueking, S. Patil, R. Pordes, A. Rana, I. Terekhov, S. Veseli, J. Yu et al., Proceedings Of The VIII International Workshop On Advanced Computing And Analysis Techniques In Physics Research, Moscow, Russia, 24 - 28 June 2002, NIM A502 (423-425), April 2003

"*MySRB & SRB - Components of a Data Grid"*, A. Rajasekar, M. Wan, R. Moore, 11th HPDC Conference, Edinburgh, Scotland, July, 2002

"A globally-distributed grid monitoring system to facilitate HPC at D0/SAM-Grid", MS Thesis, Dec. 2002, The University of Texas, Arlington; Abhishek S. Rana

"Giggle: A Framework for Constructing Scalable Replica Location Services". Ann Chervenak, Ewa Deelman, Ian Foster, Leanne Guy, Wolfgang Hoschek, Adriana Iamnitchi, Carl Kesselman, Peter Kunszt, Matei Ripeanu, Bob Schwartzkopf, Heinz Stockinger, Kurt Stockinger, Brian Tierney. Proceedings of the SC2002 Conference, Baltimore, November, 2002.

Proceedings of Computing in High Energy and Nuclear Physics Conference 2003 (CHEP'03), San Diego, CA, March 2003 (http://www-conf.slac.stanford.edu/chep03/)

Iosif Legrand, *"MonALisa: A Distributed Monitoring Service Architecture"*

Igor Terekhov, *"Grid Job and Information Management for the FNAL Run II Experiments"*

Lee Lueking, *" Dzero Regional Analysis Center Concepts and Experience"*

Rich Baker, *"A Model for Grid User Management"*

Craig Tull, *"Using CAS to Manage Role-Based VO Sub-Groups"*

Vijay Sekhri, *"Site Grid Authorization Service (SAZ) at Fermilab"*

Wensheng Deng, *"Magda - Manager for Grid-based Data"*

David Adams, *"DIAL: Distributed Interactive Analysis of Large datasets"*

Von Welch, et al., *"The Community Authorization Service: Status and Future"*

"GridFTP: Protocol Extensions to FTP for the Grid", W. Allcock, Global Grid Forum Document GFD.20, April 2003

"CA-based Trust Issues for Grid Authentication and Identity Delegation", M. Thompson, D. Olson, R. Cowles, S. Mullen, M. Helm, Global Grid Forum Document GFD.17, June 2003

Yu, D., Robertazzi, T. *"Divisible Load Scheduling for Grid Computing"*. IASTED International Conference on Parallel and Distributed Computing and Systems (PDCS 2003) (Marina del Rey, CA, Nov. 2003).

Wong, H., Yu, D., Veeravalli, B., Robertazzi, T. *"Data Intensive Grid Scheduling: Multiple Sources with Capacity Constraints"*. IASTED International Conference on Parallel and Distributed Computing and Systems (PDCS 2003) (Marina del Rey, CA, Nov. 2003).

Jagatheesan, A., R., Moore, *"Data Grid Management Systems,"* NASA / IEEE MSST2004, Twelfth NASA Goddard / Twenty-First IEEE Conference on Mass Storage Systems and Technologies, April 2004.

Moore, R., *"Preservation Environments,"* NASA / IEEE MSST2004, Twelfth NASA Goddard / Twenty-First IEEE Conference on Mass Storage Systems and Technologies, April 2004.

"DataMover: Robust Terabyte-Scale Multi-file Replication over Wide-Area Networks" Alex Sim, Junmin Gu, Arie Shoshani, Vijaya Natarajan, SSDBM'04, PPDG-43

"*Performance Analysis of the Globus Toolkit Monitoring and Discovery Service, MDS2"*, Xuehai Zhang and Jennifer M. Schopf, International Workshop on Middleware Performance (MP 2004) in conjunction with IPCCC 2004, April 2004, Phoenix, AZ

"X.509 Proxy Certificates for Dynamic Delegation," V. Welch, I. Foster, C. Kesselman, O. Mulmo, L. Pearlman, S. Tuecke, J. Gawor, S. Meder, F. Siebenlist. 3rd Annual PKI R&D Workshop, 2004.

"Performance and Scalability of a Replica Location Service", A.L. Chervenak, N. Palavalli, S. Bharathi, C. Kesselman, R.Schwartzkopf, Proceedings of the International IEEE Symposium on High Performance Distributed Computing, June 2004.

Moore, R., *"Integrating Data and Information Management,"* International Supercomputer Conference, Heidelberg, Germany, June 2004.

Moore, R., *"Digital Libraries and Data Intensive Computing,"* 2nd China Digital Library Conference, Beijing, China, September 2004

Proceedings of Computing in High Energy and Nuclear Physics Conference 2004 (CHEP'04), Interlaken, Switzerland, Sept. 2004 (http://indico.cern.ch/conferenceProgram.py?confId=0)

"A Scalable Grid User Management System for Large Virtual Organizations", G. Carcassi, et al., BNL

"Storage Resource Manager", T. Perelmutov, et al., FNAL

"ATLAS Distributed Analysis", D. Adams, BNL

"Networks and Grids for High Energy and Nuclear Physics", H. Newman, Caltech

"Grid Enabled Analysis for CMS: prototype, status and results", F. Van Lingen, Caltech, et al.

"The Clarens Grid-enabled Web Services Framework: Services and Implementation", C. Steenberg, et al., Caltech

"The Open Science Grid (OSG)", R. Pordes, FNAL, et al.

"Production mode Data-Replication framework in STAR using the HRM Grid", E. Hjort, LBNL, et al.

"Experience producing simulated events for the DZero experiment on the SAM-Grid", G.

Garzoglio, FNAL, et al.

"Mis-use Cases for the Grid", D. Skow, FNAL

"HotGrid: Graduated Access to Grid-based Science Gateways", Roy Williams, Conrad Steenberg, Julian Bunn, Proceedings of IEEE Supercomputing Conference, Pittsburgh, 2004.

"Conceptual Grid Authorization Framework and Classification", M. Lorch, B. Cowles, R. Baker, L. Gommans, P. Madsen, A. McNab, L. Ramakrishnan, K. Sankar, D. Skow, M. Thompson, Global Grid Forum Document GFD.38, Nov. 2004

A. Nishandar, "Grid-Fabric Interface For Job Management In Sam-Grid, A Distributed Data Handling And Job Management System For High Energy Physics Experiments", Thesis of Master in Computing Science, The University of Texas, Arlington, Dec. 2004.

S. Jain, "Abstracting the hetereogeneities of computational resources in the SAM-Grid to enable execution of high energy physics applications", Thesis of Master in Computing Science, The University of Texas, Arlington, Dec. 2004.

F. van Lingen, J. Bunn, I. Legrand, H. Newman, C. Steenberg, M. Thomas, A. Anjum, T. Azim, *"The Clarens Web Service Framework for Distributed Scientific Analysis in Grid Projects"* , Workshop on Web and Grid Services for Scientific Data Analysis (WAGSSDA), Oslo, June 14-17, 2005

M. Thomas,C. Steenberg, F. van Lingen, H. Newman, J. Bunn, A. Ali, R. McClatchey, A. Anjum, T. Azim,W. ur Rehman, F. Khan, J. Uk In *"JClarens: A Java Framework for Developing and Deploying Web Services for Grid Computing"*, International Conference on Web Services, Orlando, July 12-15, 2005

F. van Lingen, C. Steenberg, M. Thomas, A. Anjum, T. Azim, F. Khan, H. Newman, A. Ali, J. Bunn, I. Legrand *"Collaboration in Grid Environments using Clarens"*, 9th World Multi-Conference on Systemics, Cybernetics and Informatics , Orlando, July 2005

"RRS: Replica Registration Service for Data Grids", Arie Shoshani, Alex Sim, Kurt Stockinger, VLDB Workshop on Data Management in Grids, Trondheim, Norway, Sept. 2005

"Grid Collector: Facilitating Efficient Selective Access from Data Grids", In Proceedings of International Supercomputer Conference 2005, Heidelberg, Germany, best paper award. K. Wu, J. Gu, J. Lauret, A. Poskanzer, A. Shoshani, A. Sim, W. Zhang.

A. Nishandar, D. Levine, S. Jain, G. Garzoglio, I. Terekhov, "Extending the Cluster-Grid Interface Using Batch System Abstraction and Idealization", in Proceedings of Cluster Computing and Grid 2005 (CCGrid05), Cardiff, UK, May 2005.

A. Nishandar, D. Levine, S. Jain, G. Garzoglio, I. Terekhov, "Black Hole Effect: Detection and Mitigation of Application Failures due to Incompatible Execution Environment in Computational Grids", in Proceedings of Cluster Computing and Grid 2005 (CCGrid05), Cardiff, UK, May 2005.

G. Garzoglio, "A Globally Distributed System for Job, Data and Information Handling for High-Energy Physics"; Ph.D. Dissertation, DePaul University, Chicago; Dec 05. Ph.D. Research Proposal, Sep 04.

A. Rajendra, "Integration of the SAM-Grid Infrastructure to the DZero Data Reprocessing Effort", Thesis of Master in Computing Science, The University of Texas, Arlington, Dec. 2005.

B. Balan, "Enhancements to the SAM-Grid Infrastructure", Thesis of Master in Computing Science, The University of Texas, Arlington, Dec. 2005.

Proceedings of Computing in High Energy and Nuclear Physics Conference 2006 (Chep'06), February 2006, Mumbai, India (http://indico.cern.ch/conferenceProgram.py?confId=048)

ADAMS, David, BNL, *"DIAL: Distributed Interactive Analysis of Large Datasets"*

NORMAN, Matthew, UCSD, *"OSG-CAF - A single point of submission for CDF to the Open Science Grid"*

HJORT, Eric, LBNL, *"Data and Computational Grid decoupling in STAR – An Analysis Scenario using SRM Technology"*

HAJDU, Levente, BNL, *"Meta-configuration for dynamic resource brokering: the SUMS approach"*

RANA, Abhishek, UCSD, *"gPLAZMA (grid-aware PLuggable AuthoriZation MAnagement): Introducing RBAC (Role Based Access Control) Security in dCache"*

STEENBERG, Conrad, Caltech, *"JobMon: A Secure, Scalable, Interactive Grid Job Monitor"*

GABRIELE, Garzoglio, FNAL, *"The SAM-Grid / LCG interoperability system: a bridge between two Grids"*

T.S. Reddy, "Bridging Two Grids: The SAM-Grid / LCG integration Project", Thesis of Master in Computing Science, The University of Texas, Arlington, May 2006.

T.S. Reddy, D. Levine, G. Garzoglio, A. Baranovski, P. Mhashilkar, "Trust Model and Credential Handling for Job Forwarding in the SAM-Grid/LCG Interoperability Project", submitted to the 7th IEEE International Conference on Grid Computing (GRID 06), Barcelona, Sep. 2006.

A. Bobyshev, et al Caltech, Fermilab, "Lambda Station: On-demand Flow Based Routing for Data Intensive Grid Applications over Multitopology Networks", In proceedings of GRIDNETS 2006, San Jose, California, October 1-2, 2006