

## Particle Physics Data Grid

*PI's:* Richard Mount, SLAC, Miron Livny, Wisconsin, Harvey Newman, Caltech

*Steering Committee:* John Huth, Harvard (ATLAS), Tim Adye, RAL (BaBar), Lothar Bauerdick, FNAL (CMS),

Lee Lueking, FNAL (D0), Chip Watson, TJNAF, Jerome Lauret, BNL (STAR),

Miron Livny, Wisconsin (Condor), Jennifer Schopf, ANL (Globus), Ian Foster, ANL (Globus),

Reagan Moore, SDSC (SRB), Arie Shoshani, LBNL (SRM)

*Coordinators:* Ruth Pordes, FNAL, Doug Olson, LBNL

*Liaisons:* Paul Avery (iVDGL), Larry Price (HICB), Mike Wilde (GriPhyN), Torre Wenaus, Ian Bird (LCG)

([www.ppdg.net](http://www.ppdg.net))

### Summary

***The Particle Physics Data Grid, PPDG, has been pioneering large-scale collaboration between computer scientists and physicists who have seemingly unbounded needs for data-intensive computing integrating worldwide resources. Distributed computing for high-energy and nuclear physics experiments has become both a driver for, and an eager consumer of, computational data-grid technology. As a SciDAC project, PPDG has focused on the Grid technology advances, ranging from architecture to robustness, that can be driven by deployment and integration into the demanding end-to-end applications of the physics scientific process.***

The Particle Physics Data Grid, PPDG, is a collaboration of six physics experiments[1] (ATLAS, BaBar, CMS, D0, STAR, TJNAF) with four leading computer science projects[1] (Condor, Globus, SRB, SRM).

The PPDG collaboration, was created in 1999 reflecting the bold belief that computer scientists and physicists could work together to develop and deploy revolutionary technology that will revolutionize the way we do science. The physicists brought the drive of fundamental science, seeking answers to “What is the origin of matter?”, “Why is there much more matter than antimatter in the universe?”, and “Can we create and measure matter in the bizarre form of a quark-gluon-plasma that existed when the universe was billionths of a second old?” The physicists also brought vast data collections to be analyzed by the worldwide intellect of hundreds or thousands of scientists to answer these questions. Computer scientists brought the promise of Grid technology that could liberate and empower the scientific intellect through a seamless and powerful computing environment.

Computer scientists in PPDG acknowledge that their not-always-painless introduction to

the world of thousand-scientist collaborations is itself a benefit. In the world of particle physics, the Grid itself demands intra and intercontinental collaboration between middleware projects .

Driven by the data volume, the geographically dispersed nature of the physics collaborations, the long lifetime of the analyses process, and the number of scientists in the collaborations, a distributed computing model has evolved as the most effective means to enable the scientific productivity of the researchers and their students.

Leading Computer Scientists from the distributed computing community recognized the challenge of these computing environments as a tremendous opportunity to both develop and demonstrate data-grid technology as the next dominant form of global IT infrastructure. Several projects within the US, as well as other countries (especially The European Union), are developing and promoting grid technology. PPDG plays a unique role in stressing and facilitating early adoption of these technologies by the experiments to provide production services to their scientists so that the bugs and missing features that only show

their “ugly head” when used “at scale” and “under load” are identified and communicated back to the computer scientists.

Supported by SciDAC, DOE’s program for Scientific Discovery through Advanced Computing, PPDG is vertically integrating advances in computer science such as novel mechanisms and policies, grid middleware, experiment-specific applications and computing, storage and network resources to bring effective end-to-end capabilities to the scientist’s desk-top.

Part of the PPDG mission is to facilitate the integration effort between several groups and projects, each with their own set of timescales and driving forces. An effective approach to this integration is to identify a specific piece of middleware to be deployed in each physics collaboration that is likely to bring short-term benefits. Teams from the computer science and physics groups carry out specific development and integration tasks delivering end-to-end capabilities.

Three such accomplishments of PPDG that have been highlighted in news briefings during the first year of the project:

- Sustained transatlantic transfer of data (from D0, a major Fermilab physics experiment) authenticated with internationally trusted digital ID’s. [2]
- The first production simulation processing in a grid environment achieved in the US by US scientists preparing for the CMS experiment at the CERN Laboratory in Switzerland. [3]
- Robust coast-to-coast terabyte-scale file replication achieved by the STAR experiment studying quark-gluon-plasma physics at Brookhaven. [4]

Some of the additional accomplishments include: US scientists preparing the ATLAS experiment at CERN have integrated their data challenge application across a grid based on US middleware in the US and on European middleware in Europe; BaBar has established managed data transfer from the

SLAC laboratory in California to Lyon France using the Storage Resource Broker, SRB; the TJNAF laboratory has established remote data access as a grid web service.

In keeping with the goals of achieving end-to-end production services, PPDG recognized a missing connection with site computer-security personnel and so established a working group of the DOE laboratory site security teams to identify issues and recommend improvements to the grid security infrastructure.[5] Collaboration with DOE Science Grid has helped establish the public-key security services for the US physics grid community and trust relationship with European grid resources.

Each computer science group in PPDG has made improvements and extensions to its designs to meet needs uncovered by PPDG. These improvements are already benefiting numerous other efforts within DOE’s SciDAC program. To quote PPDG computer scientists “Almost every new feature required by Particle Physics turns out to be generic!”

PPDG collaborates on other joint activities with the NSF-funded Grid Physics Network, GriPhyN and International Virtual Data-Grid Laboratory, iVDGL projects under the umbrella name Trillium.

In the next year, all of the experiment groups in PPDG are planning to roll out grid enabled job scheduling services. Many challenging areas of work require further research and development to meet our goals: interactive science analysis services; troubleshooting end-to-end functional and performance of grid applications; and multi-organization authorization and service provision.

---

1 More about PPDG participants at [www.ppdg.net](http://www.ppdg.net).  
2 [www.ppdg.net/docs/news/news-item-20feb02.pdf](http://www.ppdg.net/docs/news/news-item-20feb02.pdf)  
3 [www.ppdg.net/docs/news/news-update-cmstestgrid-17may02.pdf](http://www.ppdg.net/docs/news/news-update-cmstestgrid-17may02.pdf)  
4 [www.ppdg.net/docs/news/news-25sep02.pdf](http://www.ppdg.net/docs/news/news-25sep02.pdf)  
5 [www.ppdg.net/pa/ppdg-pa/siteaa/](http://www.ppdg.net/pa/ppdg-pa/siteaa/)