

Particle Physics Data Grid: 2004-2006 – From Fabric to Physics

Enabling Frontier Science on a Globally Distributed Shared Grid – A Decade from Vision to Realization

Steering Committee (PI's in bold)
Richard Mount (SLAC), **Miron Livny (Wisc.)**, **Harvey Newman (Caltech)**, **Ruth Pordes (FNAL)**,
Lothar Bauerdick (FNAL), Roger Barlow (Man.), Ian Foster (ANL), Bruce Gibbard (BNL), John Huth (Harvard), Andy Kowalski (JLab), Bill Kramer (LBNL), Jerome Lauret (BNL), Wyatt Merritt (FNAL), Reagan Moore (SDSC), Dave Morrison (BNL), Greg Nawrocki (ANL), Doug Olson (LBNL), Don Petravick (FNAL), Arie Shoshani (LBNL)

PPDG is a collaboration of computer scientists with a strong record in Grid technology, physicists with leading roles in the software and network infrastructures for major high-energy and nuclear experiments, and staff at DOE production computing facilities, working together with a vision of a shared distributed computing fabric and middleware supporting collaborative data intensive science.

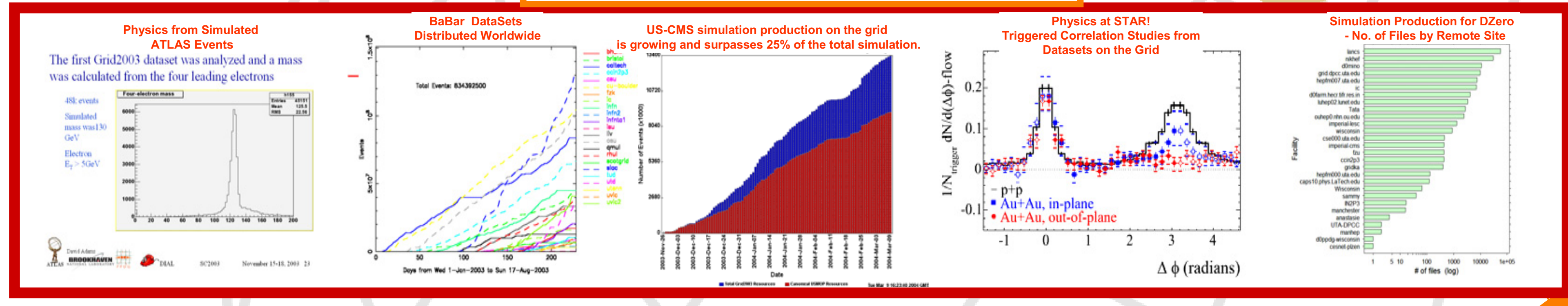
The pragmatic approach is step-wise integration of grid technology & CS concepts into experiment's end-to-end applications, as well as feedback to improving middleware, as progress towards this ultimate goal.



Contributing to Open Science Grid National Common Production Infrastructure 2004 → for Science

Standard Common Infrastructure

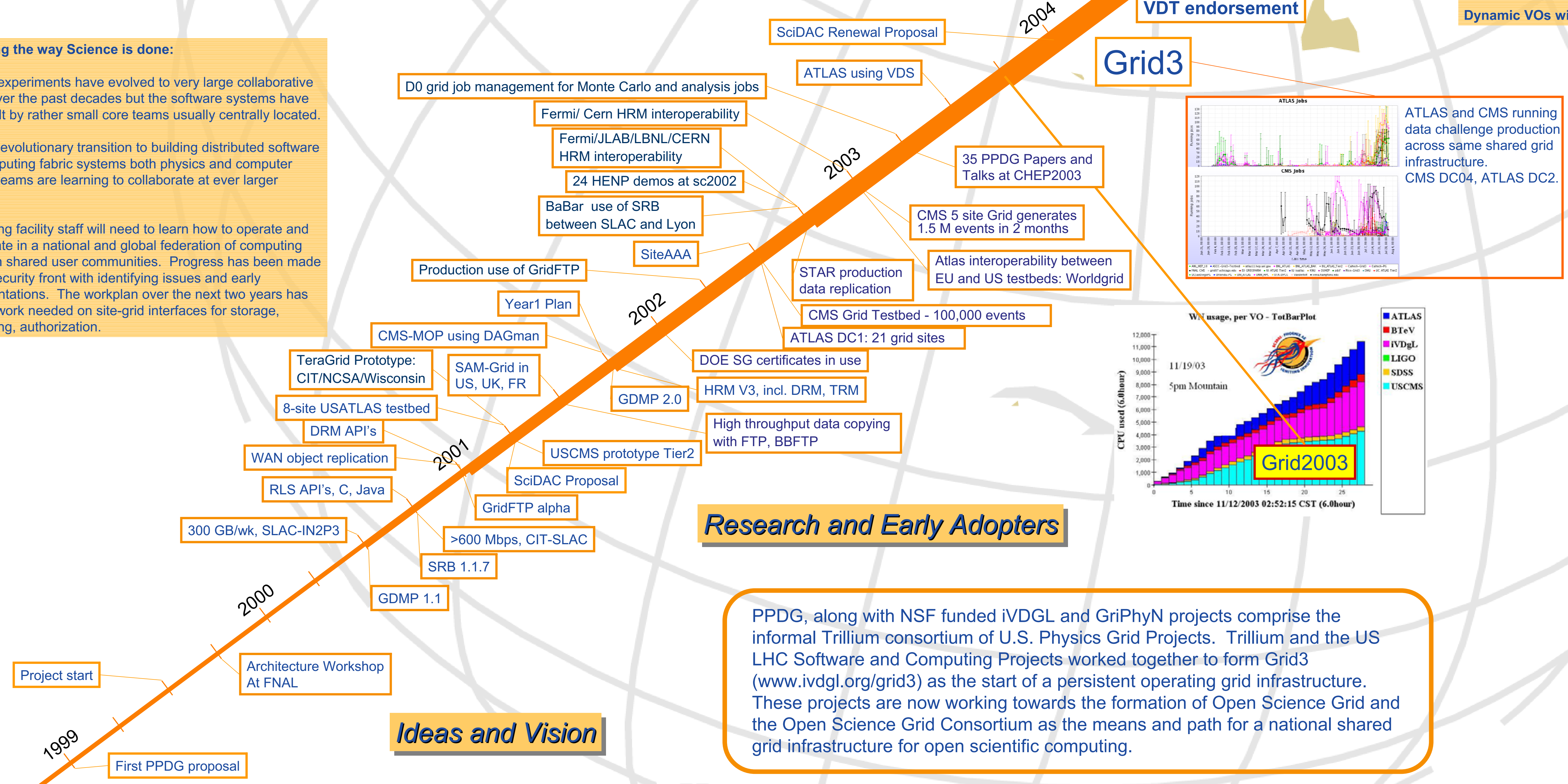
Contributing to physics in 2004



Consolidation -> Production

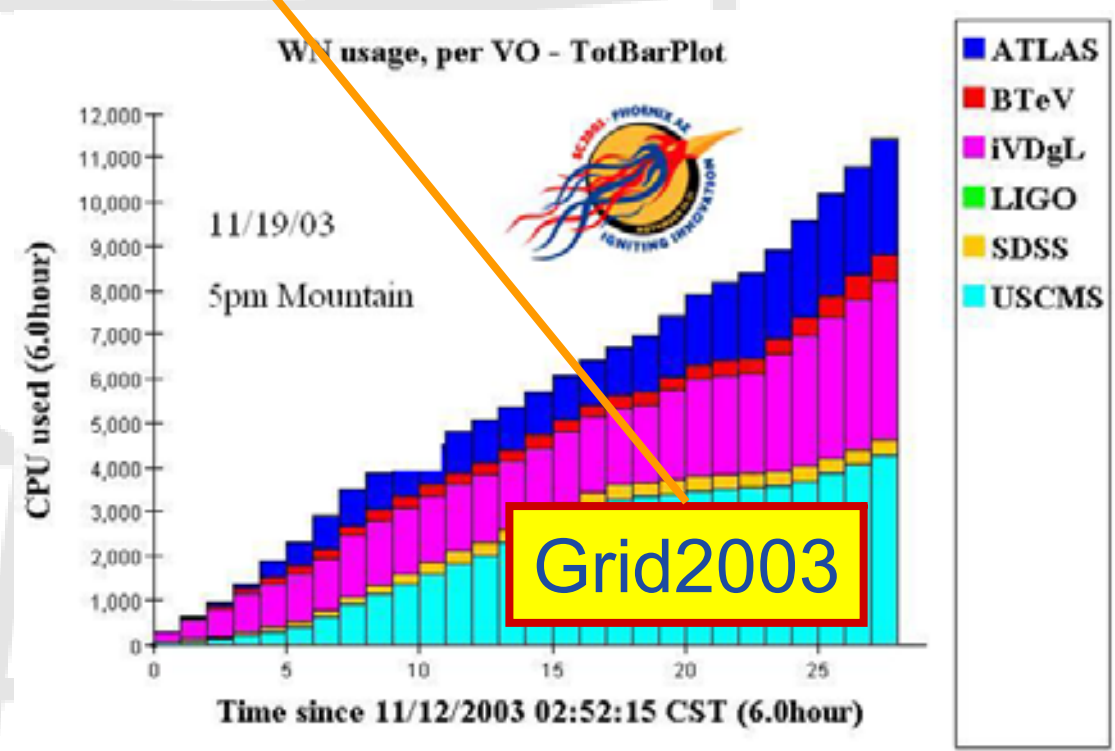
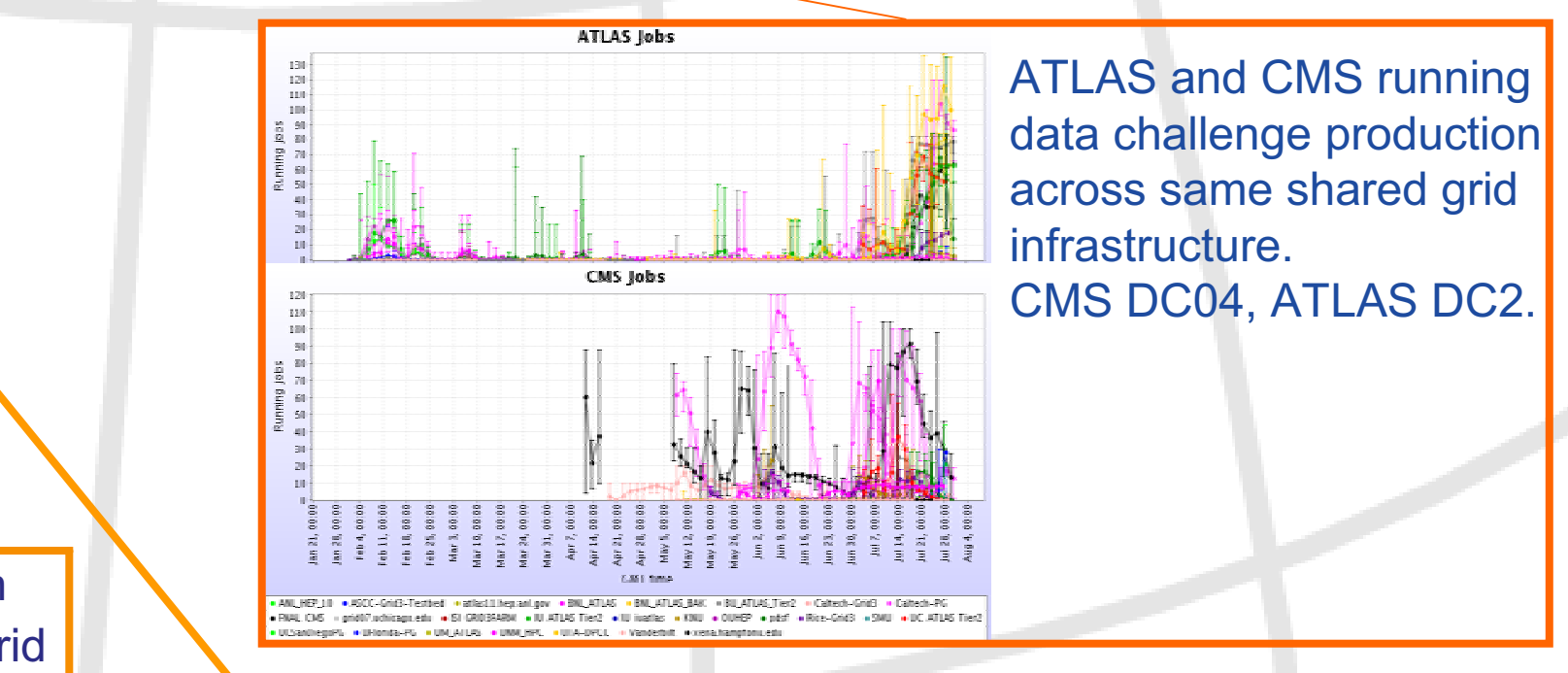
- Present DOE Laboratory Shared Facilities to the Grid Infrastructure.
- Make US LHC Resources accessible to Common Infrastructure.
- Extend Data and Job Management Models.
- Develop and Validate Operational Model.
- Iterate on Capabilities of Technologies and Services.
- Accept Contributions from Diverse Set of Participants.
- Extend the complexity of the Applications for Physics Analysis.
- Dynamic VOs with light-weight infrastructure.

Changing the way Science is done:
Physics experiments have evolved to very large collaborative efforts over the past decades but the software systems have been built by rather small core teams usually centrally located.
With the evolutionary transition to building distributed software and computing fabric systems both physics and computer science teams are learning to collaborate at ever larger scales.
Computing facility staff will need to learn how to operate and collaborate in a national and global federation of computing sites with shared user communities. Progress has been made on the security front with identifying issues and early implementations. The workplan over the next two years has charted work needed on site-grid interfaces for storage, accounting, authorization.



Research and Early Adopters

PPDG, along with NSF funded iVDGL and GriPhyN projects comprise the informal Trillium consortium of U.S. Physics Grid Projects. Trillium and the US LHC Software and Computing Projects worked together to form Grid3 (www.ivdgl.org/grid3) as the start of a persistent operating grid infrastructure. These projects are now working towards the formation of Open Science Grid and the Open Science Grid Consortium as the means and path for a national shared grid infrastructure for open scientific computing.



SAMGRid running on LCG.
STAR Distributed Analysis and Data Queries.
Babar Jobs across OSG and LCG
US LHC distributed analysis.

Changing the way Science is done:
Automating TB-scale data transport & replication
→ Less effort handling data, more effort for Science
More reliable grid-based job scheduling & execution
→ Less effort running jobs, more effort for Science
Improvements to Condor/Condor-G in PPDG benefit BLAST
→ Biology benefits from middleware improved for Physics
Enhancements to robustness & scalability of Globus, Condor in PPDG building Grid2003 let GADU run 5 times faster
→ Biology benefits from middleware improved for Physics
Computer Science & Experiment teams collaborate on building major production software systems of experiments
→ Better software systems, less effort, more time for Science
→ Computer Scientists gain deep understanding of and access to physics computational systems, new domains for computer science
ATLAS and CMS share distributed computational resources
→ More efficient use of resources, greater peak resource capacity available to each experiment
Experiments share more infrastructure software
→ More robust & functional software systems, less effort to build, more effort for Science

