

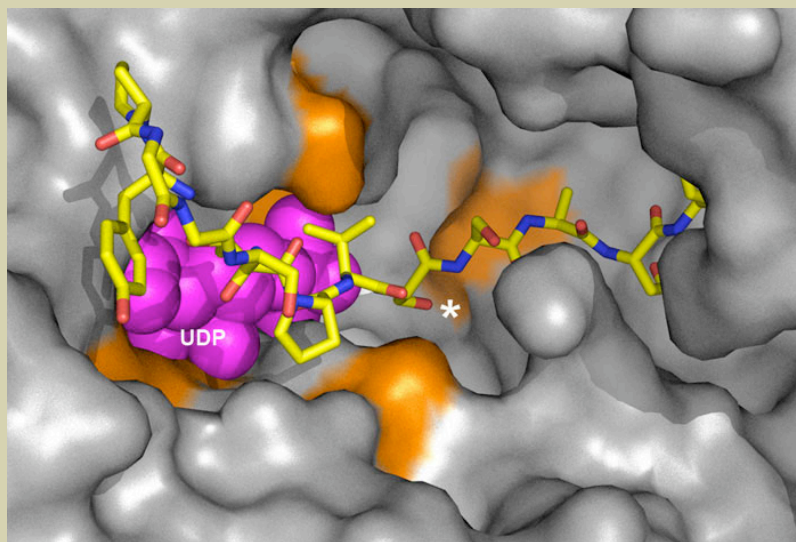


By Werner Kunz

Open Science Grid

2011

ahm2011@sbgrid.org





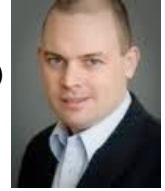
Ruth Pordes: Opening Remarks

8:30-8:55 AM, Pechet Room

9:00-12:00 PM OSG Tools and Infrastructures I

Session Chairs: Ian Stokes-Rees & Mine Altunay, Location: Pechet Room

- 9:00- 9:30: New and forthcoming hardware for cluster computing (speaker TBD)
- 9:30-10:10: Running a successful site - Steve Timm (Fermilab), Doug Johnson (Colorado)
- 10:10-10:30: Security in OSG (Mine Altunay)
- 10:30-11:00: Coffee Break
- 11:00-11:20: Running a successful VO - Rogerio Iope (GridUNESP)
- 11:20-11:40: RHIC Real time data reconstruction using Magellan cloud computing - Jan Balewski (STAR/MIT)
- 11:40-12:00: Data and application management strategies on OSG - Ian Stokes-Rees (SBGrid)



Technology Highlights: 1:45-2:30PM, Pechet

- HPC Management Suite for Cloud - Waiman Chan (IBM)
- Scale-Out NAS - Simplifying Storage and Accelerating Results - Tom DeLoria (Isilon)
- Federated Storage and Grid Storage: Adding Intelligence To Conventional File Systems - Jacob Farmer (Cambridge Computer)

2:30-5:30 PM OSG Tools and Infrastructures II

Session Chair: Dan Fraser, Location: Pechet Room

- 2:30-2:45 OSG Production: Looking to the Future - Dan Fraser (Argonne)
- 2:45-3:15 OSG Technology Plans and Direction - John Hover (Fermilab)
- 3:15-3:30 Glide-in WMS Plans and Roadmap - Burt Holzman (Fermilab)
- 3:30-3:45 Grid Operations Center - Rob Quick (IndianaU), Scott Teige (IndianaU), Kyle Gross (IndianaU)
- 3:45-4:15 Coffee Break
- 4:15-4:30 A Glide-in Factory for OSG VOs to Use - Jeff Dost (UCSD)
- 4:30-4:45 Using Pigeon Tools for Storage Discovery - Douglas Strain (Fermilab)
- 4:45-5:00 OGF standards and the OSG - Alan Sill (Texas Tech)
- 5:00-5:15 DAGMAN overview and future plans - Alain Roy (UWisconsin-Madison)
- 5:15-5:30 Globus Online and Globus Toolkit update - Steve Tuecke (UChicago/Argonne/Globus)



Security Round Table, 5:30-6:30 PM, Rm 216

Dinner at Gaslight

On Monday night there is the option of going into Boston's historic South End for a Restaurant Week dinner at the French restaurant Gaslight (gaslight560.com). The cost of a 3-course menu is included in your registration. Space is limited. Bus departing from the Best Western at 6:45PM (reservations required, please contact Michelle at ottaviano@hkl.hms.harvard.edu).

The screenshot shows the Gaslight restaurant website. It features a navigation menu with links for Home, Menu, Reservations, Group Dining, Gift Cards, Hours & Directions, Events & Donation, Contact Us, and Press. A prominent 'FREE PARKING' badge is displayed. The main content area includes a photo of the restaurant interior and a detailed description of the French brasserie cuisine, mentioning Chef de Cuisine Christopher Robins. A 'Book a Table' section is visible, showing a party size of 2, a date of 02/26/2011, and a time of 7:00 PM. The website also features social media links for Facebook and Twitter, and several award logos including 'Dining in Boston' and 'Best of Boston 2010'.

Monday 07 March 2011

09:00 - 09:15	Introduction 15' Speaker: Ken Bloom
09:15 - 09:30	Site report: Vanderbilt 15'
09:30 - 10:00	Getting ready for multicore CMS jobs 30' Speaker: Burt Holzman
10:00 - 10:15	Site report: Wisconsin 15' Speaker: Will Maier (University of Wisconsin)
10:15 - 10:30	Site report: Nebraska 15'
10:30 - 10:50	Coffee break
10:50 - 11:05	Site report: Caltech 15'
11:05 - 11:35	News from OSG 30' Speakers: Rob Quick, Alain Roy, Dan Fraser
11:35 - 11:50	Site report: Florida 15'
11:50 - 12:05	Site report: Purdue 15' Speaker: Thomas Hacker (Purdue University)
12:05 - 14:30	Lunch/vendor showcase
14:30 - 14:45	Site report: SPRACE 15'
14:45 - 15:25	Future of data access for WLCG and CMS 40' Speaker: Ian Fisk
15:25 - 15:40	Site report: UCSD 15'
15:40 - 16:00	Coffee break
16:00 - 16:15	Site report: UERJ 15'
16:15 - 16:45	xrootd demonstrator for CMS 30' Speaker: Frank Wurthwein
16:45 - 17:00	Site report: MIT 15'
17:00 - 17:15	Discussion/conclusion 15'

Tuesday 08 March 2011

09:00 - 12:00	CMS Tier 3
09:00	Recent hardware purchases 20' Speaker: all attending site admins
09:20	Care and feeding of a gatekeeper 20' Speaker: Burt Holzman (CMS)
09:40	Best security practices 20' Speaker: Will Maier (University of Wisconsin)
10:00	Discussion 30'
10:30	Coffee 20'
10:50	Bestman2 15' Speaker: Alex Sim (LBL)
11:05	Central phedex 15' Speaker: Robert Snihur (University of Nebraska-Unknown-Unknown)
11:20	How to attract jobs to your site 20' Speaker: James Letts (Department of Physics-Univ. of California at San Diego (UCSD))
11:40	Discussion 20'
14:30 - 17:30	Joint ATLAS CMS Tier 3
14:30	Cluster Management 30' Speaker: Jason Allen (Fermilab)
15:00	XROOTD Demonstrator Infrastructure 20' Speaker: Andy Hanushevsky (SLAC)
15:20	XROOTD Demonstrator for CMS 20' Speaker: Frank Wuerthwein (UCSD)
15:40	XROOTD Demonstrator for ATLAS 20' Speaker: Charles George Waldman (University of Chicago)
16:00	Coffee 20'
16:20	ATLAS Tier 3 Analysis Experience 25' Speaker: Stephane Willocq (University of Massachusetts)
16:45	Future directions for CMS Tier 3 25' Speaker: Ian Fisk (Fermi National Accelerator Laboratory (FNAL))
17:10	Discussion 20'
19:00 - 21:00	Dinner (Elements Cafe, Sponsored by Dell)

US ATLAS Facility Meeting at OSG All Hands

chaired by Michael Ernst (BNL, ATLAS) , Robert GARDNER (UNIVERSITY OF CHICAGO) , Doug Benjamin (Duke University) , Rikutaro Yoshida (Argonne National Laboratory)

from Monday 07 March 2011 at 03:30 to Tuesday 08 March 2011 at 18:00 (US/Eastern)
at Harvard Medical School

Description The next in our series of face-to-face workshops focused on topics related to the USATLAS distributed facility computing operations and integration, including Tier 1/2/3 centers.

Meeting website, <http://ahm.sbgrid.org/>

[Go to day](#) ▾

Monday 07 March 2011

- 09:00 - 09:30 **ATLAS Requirements and the US ATLAS Computing Facility** 30'
Speaker: Michael Ernst (BNL, ATLAS)
- 09:30 - 09:50 **Facilities Integration Topics** 20'
Speaker: Robert W Gardner (University of Chicago)
- 09:50 - 10:30 **ATLAS Roadmap for Data Management and Caching** 40'
Speaker: Kaushik De (UT-Arlington)
- 10:30 - 10:50 **Coffee**
- 10:50 - 11:10 **CVMFS at Tier 2** 20'
Speaker: Sarah Williams (Indiana University)
- 11:10 - 11:40 **Data Distribution Performance** 30'
Speaker: Hironori Ito
- 12:00 - 13:30 **Lunch**
- 14:00 - 14:30 **Physics Analysis Experiences, Feedback** 30'
Speaker: Kevin Black (Harvard University)
- 14:30 - 14:50 **Analysis Metrics** 20'
Speaker: Sergey Panitkin (Department of Physics - Brookhaven National Laboratory (BNL))
- 14:50 - 15:10 **Athena multi-core and the Facility** 20'
Speaker: Paolo Calafiura (Lawrence Berkeley National Lab. (LBNL))
- 15:10 - 15:40 **USATLAS Networking: Status and Discussion** 30'
Speaker: Shawn McKee (University of Michigan ATLAS Group)
- 15:40 - 16:00 **Coffee**
- 16:00 - 17:00 **Discussion** 1h00'

ATLAS

Tuesday 08 March 2011

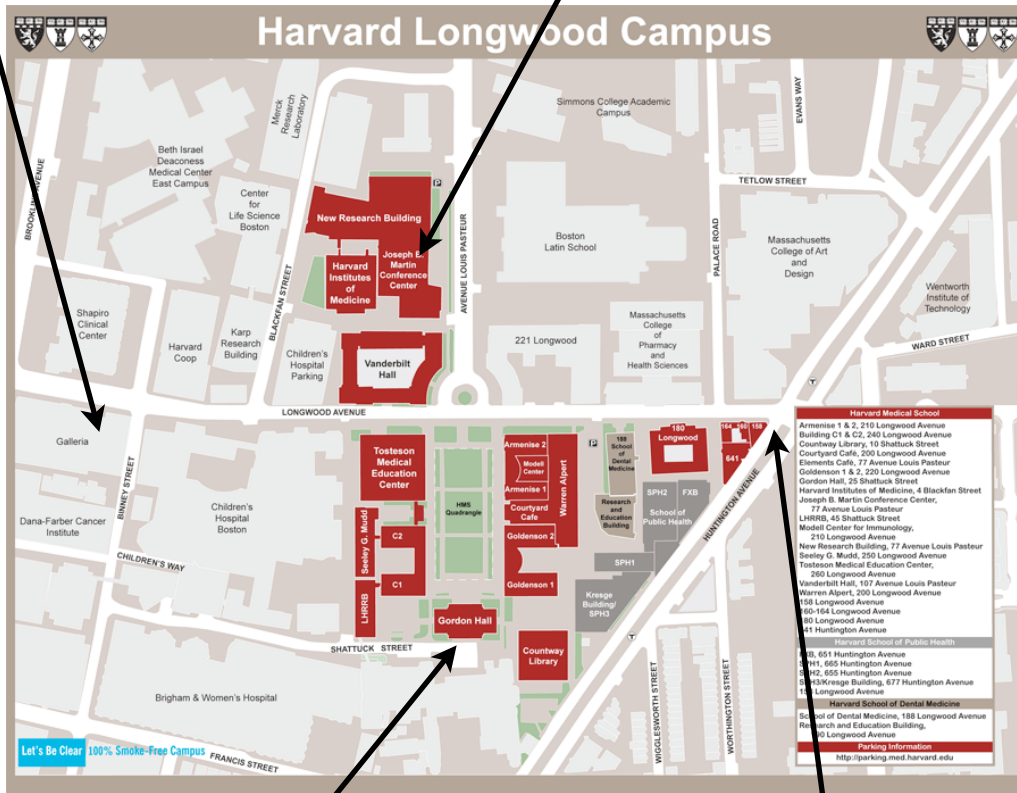
- 09:00 - 09:15 **US Atlas Tier 3 Introduction/Status** 15'
Speaker: Doug Benjamin (Duke University)
- 09:15 - 09:45 **Tier 3 Panda status and plans** 30'
Speaker: Alden Stradling (UT Arlington)
- 09:45 - 10:15 **Proof Analysis Clusters** 30'
Speaker: Sergey Panitkin (Department of Physics - Brookhaven National Laboratory (BNL))
- 10:15 - 10:45 **Coffee/Tea Break**
- 10:45 - 11:15 **XRootd federation** 30'
Speaker: Charles George Waldman (University of Chicago)
- 11:15 - 11:45 **Xrootd Federation discussion** 30'
discussion on xrootd federation issues between Tier 1, Tier 2 and Tier3 sites
Speakers: Doug Benjamin (Duke University) , Andrew Hanushevsky (STANFORD LINEAR ACCELERATOR CENTER) , Charles George Waldman (University of Chicago) , Wei Yang (SLAC) , Robert GARDNER (UNIVERSITY OF CHICAGO) , Hironori Ito
- 14:30 - 17:30 **Joint ATLAS CMS Tier 3**
- 14:30 **Cluster Management** 30'
Speaker: Jason Allen (Fermilab)
- 15:00 **XROOTD Demonstrator Infrastructure** 20'
Speaker: Andy Hanushevsky (SLAC)
- 15:20 **XROOTD Demonstrator for CMS** 20'
Speaker: Frank Wuerthwein (UCSD)
- 15:40 **XROOTD Demonstrator for ATLAS** 20'
Speaker: Charles George Waldman (University of Chicago)
- 16:00 **Coffee** 20'
- 16:20 **ATLAS Tier 3 Analysis Experience** 25'
Speaker: Stephane Willocq (University of Massachusetts)
- 16:45 **Future directions for CMS Tier 3** 25'
Speaker: Ian Fisk (Fermi National Accelerator Laboratory (FNAL))
- 17:10 **Discussion** 20'
- 19:00 - 21:00 **Dinner** (Elements Cafe, Sponsored by Dell)



The Best Western



Conference Center
Mon-Wed



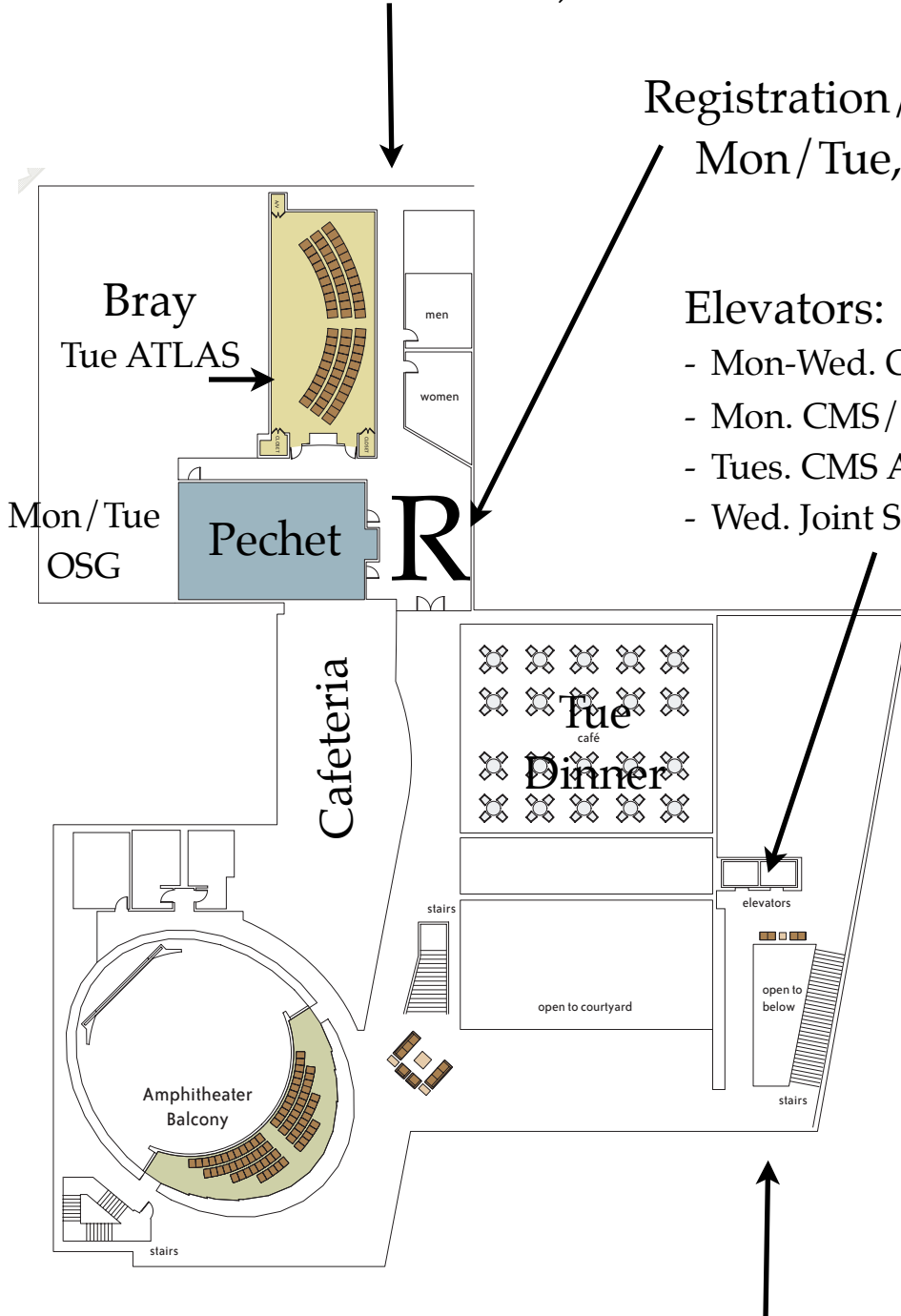
Gordon Hall
Thu AM



Longwood Medical
Green Line E - Stop
(0.3 mile walk to the Best
Western)

Main Floor

Blackfan Circle Entrance
(shortcut to the Best Western)



Registration/Breakfast
Mon/Tue, 8-9AM

Elevators:

- Mon-Wed. OSG Lunch Sitting Area: 2nd Fl
- Mon. CMS/ATLAS: Rm 214/216.
- Tues. CMS AM Session, Rm 216.
- Wed. Joint Session, Third Floor.

Louis Pasteur Ave Entrance



8:00 - 8:30AM Breakfast & Registration



Molecular Movies: Exploring Hollywood's Tools for Biovisualization

Gael McGill, 8:30-8:55 AM, Pechet Room

9:00-12:00PM Focus on High Throughput Parallel Computing (HTPC)

Session Chair: [John McGee](#), Pechet Room

- 9:00– 9:10 HTPC session intro and objectives - John McGee (UNC-CH)
- 9:10– 9:30 HTPC Engagements - Steven Cox (RENCI)
- 9:30–10:00 Campus Grid Technology - Derek Weitzel (UNebraska-Lincoln)
- 10:00–10:30 On the use of VM for OSG VOs - Lance Stout (Clemson)
- 10:30–11:00 Coffee Break
- 11:00–11:15 Where do we go from here? - Dan Fraser (Argonne)
- 11:15–12:00 Panel Discussion (McGee, Weitzel, Cox, Stout, Fraser, Thain)



Vendors Showcase Session: 12:00-2:30 PM, Rm 216

Coffee and snacks will be served



Science Highlights Session: Physics at CMS Tier 3

Wei Li, 2:00-2:25 PM, Pechet Room

2:30-5:30PM Technology Tuesday II - Focus on Grid Computing.

Session Chair: [Igor Sfiligoi](#), Location: Pechet Room

- 2:30-2:50 Potential Data Access Architectures using Xrootd - Andy Hanushevsky (SLAC)
- 2:50-3:05 Experience with Wide Area Lustre - Dave Dykstra (Fermilab)
- 3:05-3:20 The Power of HTTP Proxy Caches - Dave Dykstra (Fermilab)
- 3:20-3:40 Using Pegasus 3.0 for Data-based workflows on the OSG - Mats Rynge (ISI/USC)
- 3:40-4:00 Co-scheduling of network and storage with StorNet - Alex Sim (Lawrence Berkeley)
- 4:00-4:30 Coffee Break
- 4:30-4:50 The DYNES Architecture LHC Data Movement - Shawn McKee (UMich)
- 4:50-5:10 Experiences from Running the PanDA Pilot Factories - Xin Zhao (BNL)
- 5:10-5:30 How Monalisa is used in LHC experiments - Michael Thomas (CALTECH)



Dinner at the Conference Center, 7PM

Sponsored by Dell



8:00 - 8:30AM Breakfast - sponsored by Isilon

OSG Virtual Organizations - Status and Outlook

8:30-10:30AM

ALICE, Belle 2, CMS, CDF, DZero, Engage, HCC, LIGO, SBGrid, USATLAS

Coffee Break - sponsored by Isilon



Science Highlights Session: LHC

Prof John Huth, 11-11:40AM



OSG Status

Ruth Pordes, Miron Livny, Paul Avery

11:40-12:20

2:00-2:40 Global Cyberinfrastructure Panel

S. Newhouse (EGI), P.Papadopoulos (SDSC), J. Caballero (OSG Outreach to Americas), W. Gentzsch (DEISA),

2:40-3:30 US Cyberinfrastructure Panel

M. Livny (OSG), J. Towns (Teragrid), P. Papadopoulos (SDSC), P. Maechling (SCEC), R. Eigenmann (NEES)

Coffee Break - Sponsored by Cambridge Computers

4:00-4:30 HTPC Technology Panel

J McGee (UNC-CH), D Fraser (Argonne), Ray Scott (Carnegie Mellon)

4:30-5:00 OSG Technology Panel

Roy (UWisconsin-Madison), I. Sfiligoi (UCSD), J. Hover (Brookhaven), I. Stokes-Rees (HMS)

5:00-5:30 Resource Sharing Panel

F. Wuerthwein (UCSD), R. Gardner (UChicago), D. Fraser (Argonne)



Closing Remarks

Piotr Sliz

Reception at the Conference Center,
5:30-7PM, Sponsored by IBM

Questions for panelists:

US CI:

1. There is a hierarchy of cyberinfrastructures in place in the US ranging from campuses, laboratories, and a plethora of grids. Will these merger or diverge in common infrastructure in the days ahead
2. How do standards, APIs, and grid software "libraries" or "SDKs" fit into the US CI picture?

Global CI:

1. What are some of the most important developments in EGI that i) OSG could benefit from; and ii) that individual researchers or VOs could benefit from, regardless of OSG-wide adoption?
2. What are the emerging issues and emerging technologies to deal with the vastly different policies regarding authentication and authorization and the secure utilization of computing resources and the data being migrated around the globe in an effort to provide a global cyberinfrastructure?

HTPC:

1. What types of applications will benefit from the increasing number of multicore processors found on a worker node now and in the future. Will there be a sweet spot in terms of the number of cores (due to the sharing of memory, etc) on these same worker nodes based on the current suite of applications that could be ported to HTPC.
2. What are some successes of the existing HTPC work that have enabled jobs not possible with the standard, single-core job slots?

OSG Tech Panel:

1. What is the biggest challenge in OSG technology today? in five years?
2. What would be a single software/VDT technology that would make your computing on OSG significantly easier. Can you briefly describe your prototype.

Resource Sharing:

1. How do you see the future of resource sharing evolving as the big resource providers become locked into full scale production running? What will come along to fill in the gaps?
2. Currently it is a big deal to install the OSG & CMS software on an existing professionally-managed university cluster. Some universities have done it, but others, such as my own, cannot because of the large overhead of the software and security worries. Is there any possibility all this could be bundled in a way that it looks just like another package to be installed (like mathematica) so that we could make use of large university computing resources?

Talk with the Experts session at the OSG all hands workshop

Thursday, 9-12PM, Gordon Hall, The Fenway Room

10 min walk from The Best Western -- see map on page 2

The "talk with the experts" session is an open, loosely structured interactive session where OSG experts are available to guide and support users (scientists and system administrators) by answering questions or providing one-on-one coaching.

Check <https://twiki.grid.iu.edu/bin/view/SiteCoordination/ExpertsAtAllHands2011> for pointers, reference material, tutorials and information that you may be useful to check in advance and also for the coordinates to join remotely if you are not in Boston.

To help our planning, submit your questions in advance by sending an email to marco@hep.uchicago.edu

Experts include:

- Miriam Boon
- Robert Engel
- Gabriele Garzoglio
- Andrew Hanushevsky
- Scot Kronenfeld
- Tanya Levshina
- Marco Mambelli
- Alain Roy
- Igor Sfiligoi
- Doug Strain
- Suchandra Thapa
- Steve Timm
- Jim Weichel
- Jason Zurawski

Breakfast, 9AM

Bag Lunch will be served at 11:30AM



Council Meeting



- *Council Meeting I: Thursday 2-6PM, The Best Western Room 502*
- *Dinner at the Harvard Faculty Club: bus departing from the Best Western at 6:45PM, Thursday (reservations required, please contact Michelle at ottaviano@hkl.hms.harvard.edu).*



- *Breakfast, Friday, 8:30AM, Best Western, The Fenway Room*
- *Council Meeting II: Friday, 8:30-12PM, The Best Western, The Fenway Room*

Computing for Clinical, Pharmaceutical and Basic Life Science Research

CLSB Conference Room (3rd floor)

Thursday, 9-2PM

- 9:00-9:45AM Structural Biology Computing:
 - Compute and data management strategies for grid deployment of high throughput protein structure studies - Ian Stokes-Rees (SBGrid Virtual Organization):
 - The WeNMR project: building and operating a worldwide e-Infrastructure for NMR and structural biology - Marco Verlatto (WeNMR project, Sezione di Padova, Italy)
 - Centrally managed scientific software support consortium: benefits and challenges - Ben Eisenbraun (SBGrid)
- 9:45-10:25AM US Cyberinfrastructure Technologies
 - Open Science Grid - Miron Livny (U. Wisconsin, OSG)
 - TeraGrid - John Towns (NCSA)
- 10:45-12:30PM Research Computing Highlights (10:45 - 12:30) 5 x ~ 20' talks
 - TBD
- 12:30-12:45PM Lunch Served
- 12:45-1:35PM Boston Infrastructure Highlights
 - Orchestra Cluster at Harvard Medical School - Mark Komarinski (RITG)
 - Odyssey Cluster at Harvard University - James Cuff (FAS)

"Airport" stop on Blue Line



Copley Plaza



Green Line

Longwood Stop
Green Line - E

Government Center
Green/Blue Lines

Subway: The T

The subway is commonly called the "T" and operates from 5:30 a.m. to 12:30 a.m.; rides are \$1.70 when purchased in advance, or \$2.00 when purchased on the train. The four subway lines are called the Blue, Green (B, - E), Red, and Orange lines. In addition to the T, there are also many bus lines serving the Boston area. Maps and schedules are available at South Station and other major station stops. For maps and more information, see the MBTA (Massachusetts Bay Transportation Authority) website: www.mbta.com

Public Transit From Logan Airport

Take the SL1 Silver Line to South Station (Departs from Terminal A). Exit the Silver Line at South Station Stop. Take the Red Line from South Station (Inbound-Alewife). Exit at the Park Street Stop Change to the Green Line (E Line - Heath Street on Track 2). Exit at the Longwood Medical Area Stop. Cross Huntington Ave to Longwood Ave on foot. Walk 1 block to Avenue Louis Pasteur.



Open Science Grid

June 26-30, 2011

University of Wisconsin-Madison

2011

OSG SUMMER SCHOOL

Harness the power of distributed computing

- Use high-throughput computing and the Open Science Grid
- Run thousands of jobs and handle terabytes of data
- Learn by doing—lots of hands-on activities
- Taught by faculty & staff who work with distributed computing daily

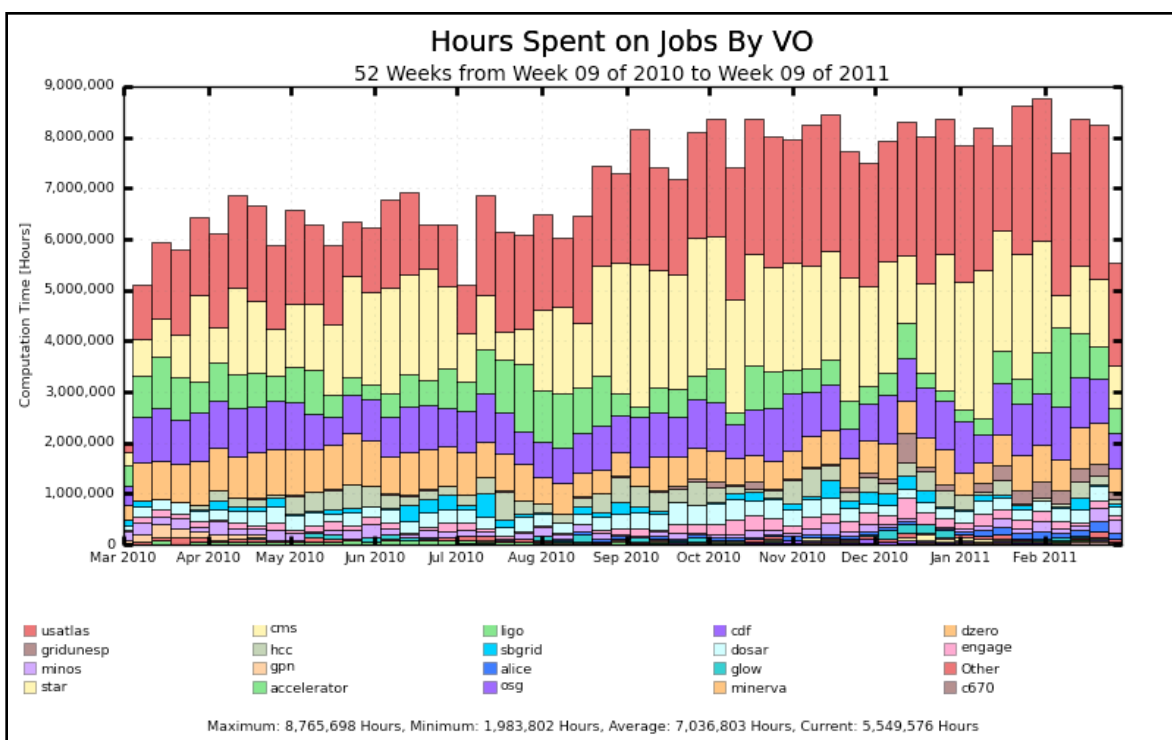
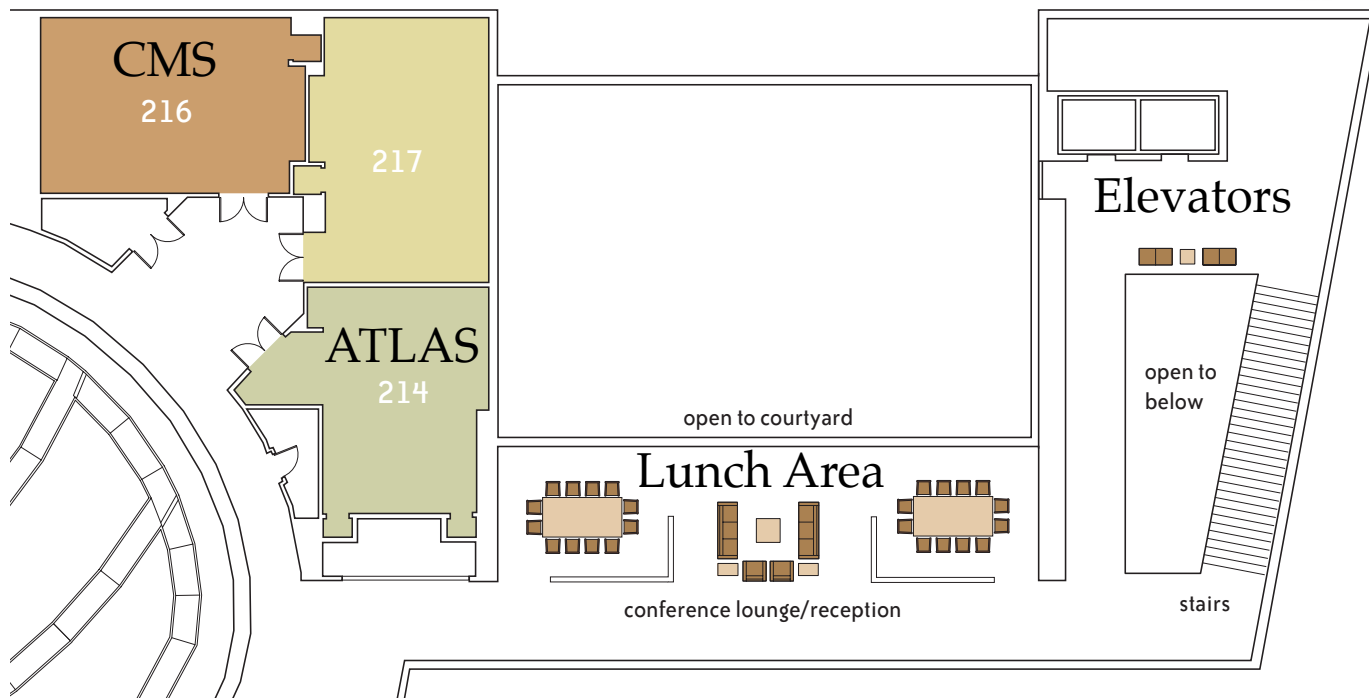
Ideal for graduate students (also faculty & staff) whose research involves large-scale computing:

Physics • Biology • Chemistry • Meteorology • Computer Science • & others!

More information and to apply (by April 1st):

www.opensciencegrid.org/GridSchool

Second Floor - Monday



Registered Participants

Abbott, Brad - University of Oklahoma
Allen, Jason- Fermilab
Almes, Guy- Texas A&M University
Altunay, Miney - Fermilab
Ananthan, Balamurali- Tech-X Corporation
Attebury, Garhan - University of Nebraska-Lincoln
Avery, Paul - University of Florida
Bauerdick, Lothar - Fermilab
Bean, Ralph - Rochester Institute of Technology
Benjamin, Douglas - Duke University
Bernick, David - Cambridge Computer
Blackburn, Kent - Caltech
Bloom, Ken - University of Nebraska-Lincoln
Buterbaugh, Kevin - Vanderbilt University
Caballero, Jose - Brookhaven National Laboratory
Cartwright, Tim - University of Wisconsin
Castro, Harold - Universidad de los Andes
Chan, Waiman - IBM
Clark, Steven - Purdue University
Cox, Steven
Danuser, Gaudenz - Harvard Medical School
Daudert, Britta - Caltech
Deelman, Ewa - USC Information Sciences Institute
Doherty, Peter - Harvard Medical School
Dost, Jeffrey - UCSD
Dykstra, Dave - Fermilab
Ernst, Michael - BNL
Farmer, Jacob - Cambridge Computer
Fisk, Ian - Fermilab
Fowler, Jack - Brown University
Fraser, Dan - Argonne National Laboratory
Fu, Yu - University of Florida
Gardner, Rob - University of Chicago
Garzoglio, Gabriele - Fermilab
Gentzsch, Wolfgang - DEISA
Gray, Paul Stephen - Dell, Inc.
Gross, Kyle - Indiana University/OSG
Hacker, Thomas - Purdue University
Hanushevsky, Andrew - SLAC National Accelerator
Laboratory
Holzman, Burt - Fermilab
Hover, John - Brookhaven National Lab
Hoyt, Josh - Galois, Inc.
Iope, Rogerio - Sao Paulo State University
Johnson, Douglas - University of Colorado
Johnson, Glenn - University of Iowa
Kriegel, Gary - Dell Inc
Kronenfeld, Scot - University of Wisconsin
Lauret, Jerome - Brookhaven National Laboratory
Letts, James - UCSD
Levshina, Tanya - Fermilab
Li, Wei - MIT
Livny, Miron - University of Wisconsin-Madison
Luehring, Frederick - Indiana University
Maechling, Philip - Southern California Earthquake Center
Maier, Will - University of Wisconsin
Mambelli, Marco - University of Chicago
McGee, John - UNC-CH
McGuigan, Patrick - UTA
McKee, Shawn - University of Michigan
Milroy, Daniel - University of Colorado
Moore, Michael - Clemson University
Moore, Richard - U California San Diego
Mora, John - Vanderbilt University
Mount, Richard - SLAC
Nahn, Jennifer - Fermilab
Napier, Austin - Tufts University
Navarro, John-Paul - Univ of Chicago/Argonne
O'Donovan, Daniel J. - Harvard Medical School
Olson, Douglas - LBNL
Panitkin, Sergey - Brookhaven National Lab
Papadopoulos, Philip - University of California, San Diego
Paugh, Jim - Advanced Clustering Technologies
Pordes, Ruth - Fermilab
Porter, Jeff - LBNL
Potekhin, Maxim - Brookhaven National Laboratory
Potoczny-Jones, Isaac - Galois
Quick, Rob - Indiana University
Riley, Daniel - Cornell University
Rizwan, Papia - University of Texas at Brownsville
Robinson, John-Paul - University of Alabama at
Birmingham
Ross, Barrett - Southern Methodist University
Roy, Alain - University of Wisconsin-Madison
Rynge, Mats - Information Sciences Institute
Scott, Ray - Carnegie Mellon University
Sehgal, Chander - Fermilab
Severini, Horst - University of Oklahoma
Sfiligoi, Igor - UCSD
Sheldon, Paul - Vanderbilt University
Sill, Alan - Texas Tech University
Sim, Alex - Lawrence Berkeley National Laboratory
Sliz, Piotr - Harvard Medical School
Snider, Rick - Fermilab
Snihur, Rob - UNL/Fermilab
Snow, Joel - Langston University
Socolovsky, Eduardo - Center for Biotechnology, Norfolk
State University
Sosebee, Mark - University of TX Arlington
Squires, Daniel - University of Iowa HEP Group
Squires, Michael - UC Davis
Stokes-Rees, Ian - Harvard Medical School

Stout, Lance - Clemson University
Stradling, Alden - UT Arlington
Strain, Douglas - Fermilab
Strossman, William - Univ. of California, Riverside
Suh, In-Saeng - University of Notre Dame
Swanson, David - University of Nebraska-Lincoln
Teige, Scott - Indiana University
Testa, Lucy - Isilon Systems
Thapa, Suchandra - University of Chicago
Thomas, Michael - California Institute of Technology
Timm, Steven - Fermilab
Tonjes, Marguerite - University of Maryland
Towns, John - National Center for Supercomputing Applications (NCSA)
Tuecke, Steven - UChicago / Argonne / Globus
Verlato, Marco - INFN
Waldman, Charles - University of Chicago
Wang, Nanbor - Tech-X Corporation

Wang, Shaowen - University of Illinois at Urbana-Champaign / National Center for Supercomputing Applications
Weichel, Jim - Fermilab/University of Florida
Weitzel, Derek - University of Nebraska - Lincoln
Wenaus, Torre - BNL
White, Victoria - Fermilab
Williams, Sarah - Indiana University
Wuerthwein, Frank - UCSD
Yang, Wei - SLAC
Zhao, Xin - Brookhaven National Lab
Zurawski, Jason - Internet2
Padmanabhan, Anand - University of Illinois at Urbana-Champaign / National Center for Supercomputing Applications
Eisenbraun, Ben - Harvard Medical School
Komarinski, Mark - Harvard Medical School
McCaulay, Scott - Indiana University
Piilonen, Leo - Virginia Tech

Speakers and Session Chairs

Paul Avery is professor of physics at the University of Florida working in the areas of experimental high energy physics and advanced computing. Avery is a Fellow of the American Physical Society and collaborates on the CMS experiment currently taking data at CERN's Large Hadron Collider. He has been principal investigator and co-PI on a number of national distributed computing projects, including Open Science Grid where he serves as co-Chair of the OSG Council.

Jose Caballero is a physicist and electronic engineer working in the Physics Applications Software group at Brookhaven National Laboratory. He started his research career in the CMS experiment, and is now part of the OSG staff and a member of the ATLAS collaboration. As an ATLAS member he develops software for the PanDA Workload Management System; as an OSG member he is the coordinator of outreach for South America.

Chander Sehgal is the project manager for the Open Science Grid and is located at Fermilab in Batavia, Illinois. He has broad experience in managing very large software projects at Bell Labs where he led the development and deployment of high-availability systems. Before joining Fermilab, he was co-owner and CFO of a start-up wireless internet company. Sehgal's areas of interest include program and project management of technology solutions; business case development and financial management of large projects; integration testing of complex technology solutions; and, process design and improvement. He has an MS in Computer Engineering from Stanford University.

Rob Gardner is a senior research associate in the Enrico Fermi Institute and senior fellow at the Computation Institute at the University of Chicago. Gardner has coordinated the OSG Integration and Sites area from the beginnings of OSG and now works on OSG metrics analytics. Rob is an ATLAS physicist and the principal investigator for the US ATLAS Midwest Tier 2 Center, a computing facility for ATLAS on the OSG, and as well manages the US ATLAS distributed facility integration program.

Dr. Wolfgang Gentzsch is a strategic consultant for HPC, Grid, and Cloud infrastructures, for governments, research, and industry, world-wide. He is Director At Large of the OGF Open Grid and Cloud Forum, OGF for e-Infrastructure Standards and an advisor to the Distributed European Infrastructure for Supercomputing Applications (DEISA). From 2004–2008, Wolfgang was a member of PCAST, the US President's Council of Advisors for Science and Technology. He directed the \$150 Mio German D-Grid Initiative, and was a professor for computer science at Duke University in Durham and NC State in Raleigh, North Carolina, and at the University of Applied Sciences in Regensburg, Germany.

John Hover is Grid Group Leader at the RHIC / ATLAS Computing Facility at Brookhaven National Lab. He is a software engineer and systems administrator who has worked in academic and scientific computing for 15 years. Currently he contributes to the OSG Blueprint group and is the lead developer for GUMS (the OSG site authorization tool).

Miron Livny is a senior researcher and professor specializing in distributed computing at the University of Wisconsin–Madison. Livny has been a professor of computer science at Wisconsin since 1983, where he leads the Condor high-throughput computing system project. He is also a principal investigator and currently the technical director for the Open Science Grid project.

Philip Maechling is the Information Technology Architect for the Southern California Earthquake Center (SCEC) at the University of Southern California. He has been managing SCEC's research computing since 2002, including the SCEC Community Modeling Environment (SCEC/CME), an earthquake system project that develops and validates physics-based computational models of earthquake processes and assembles those models into earthquake forecasting systems, and the SCEC Collaboratory for the Study of Earthquake Predictability (CSEP), an international earthquake forecast testing facility.

John McGee is the Director of Cyberinfrastructure for the Renaissance Computing Institute at UNC Chapel Hill. McGee leads the Engagement Program, a satellite project of the OSG helping users from a broad range of disciplines to leverage the power of the Open Science Grid and scale their computational science to the national level.

Gaël McGill is the founder & CEO of Digizyme, Inc. a firm dedicated to the visualization and communication of science through advanced technology applications, and the creator of the online portal molecularmovies.org and the Molecular Maya software toolkit. He is also the Digital Media Director for E.O. Wilson's 'Life on Earth' next-generation digital biology textbook, and teaches scientific visualization at Harvard Medical School where he carries out research on the impact of biovisualization in student learning.

Steven Newhouse is Director of EGI.eu and Project Director of the EGI-InSPIRE project. EGI.eu is a new organisation established last year in Amsterdam with a staff of 20, and with around 20 staff within the community, that coordinates the European Grid Infrastructure on behalf of its stakeholders - national and domain specific resource providers and the user communities they support. Steven has a background in HPC and computational numerics, and has been active in the grid community since 2001. He is chairman of the OGF Board.

Alain Roy is a researcher with the Condor Project at the University of Wisconsin-Madison and the Software Coordinator for the Open Science Grid. He is a delightful fellow, full of smiles and laughter.

Igor Sfiligoi is a software developer and researcher at the University of California San Diego. Sfiligoi is currently the leader of the OSG Scalability, Reliability and Usability area, leader of the OSG glidein factory operations, member of the OSG security team and glideinWMS development team. He has been involved in distributed computing, mostly within High Energy Physics experiments, since his graduation in 1997.

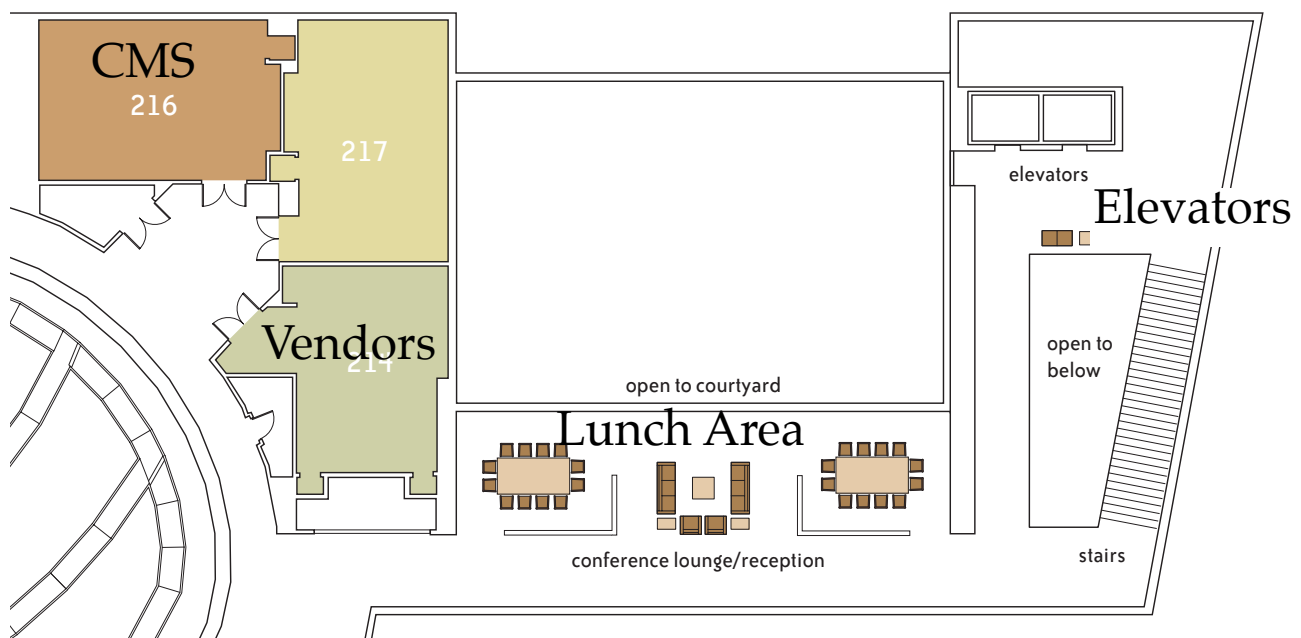
Piotr Sliz is an assistant professor of pediatrics at Harvard Medical School. His research focuses on structural biology, and structural biology computing. Sliz is the principal investigator of the SBGrid Consortium and the NSF Research Coordination Network for Structural Biology Computing. He is also the coordinator of the SBGrid Virtual Organization and serves as a member of the OSG Council. Dr Sliz is the host of the OSG AHM2011 meeting.

Ian Stokes-Rees, Ph.D. is an expert in large scale grid architectures. He has established a web-based computational environment for the SBGrid Consortium that executes large workflows using the national Open Science Grid cyberinfrastructure. His specific interests are in RESTful models for generic computational grids. He is also an expert consultant in the ETSI Grid Specialist Task Force, commissioned by the European Union, where he advises on the state of the art in grid standards and interoperability.

John Towns is the director of the Persistent Infrastructure Directorate at NCSA/University of Illinois and is also the Chair of the TeraGrid Forum. His background is in computational astrophysics utilizing a variety of computational architectures with a focus on application performance analysis. Towns is the principal investigator for several HPC system deployment and operation awards and for a recent award for evaluating emerging technologies for distributed cyberinfrastructures.

Frank Wuerthwein is a professor of physics at UCSD. His physics research focuses on searching for new phenomena, especially dark matter, with the CMS detector at the Large Hadron Collider. He is responsible for data analysis computing of the CMS experiment, and is one of the two Application Coordinators of the Open Science Grid. His group at UCSD operates a CMS Tier-2 center, as well as the glideinWMS instance used by several OSG VOs

Second Floor - Tuesday



Grid scatters light on protein structures

FEATURE | DECEMBER 15, 2010 | BY [MIRIAM BOON](#)

Structural biologists who have spent months or even years trying to determine the structure of intractable proteins may be able to find solutions overnight, thanks to the power of grid computing.

Many advances in medicine and biology depend on understanding how proteins interact with each other and other factors such as drugs, RNA, or DNA. But to do that, researchers must determine how proteins are shaped. It isn't enough to know a protein's sequence; the curly strands of a protein can bend and twist in any number of directions, changing the way it will interact with its environment.

To solve the structure of an unknown protein, structural biologists begin by crystallizing the protein. Then they visit a synchrotron, where they place their crystal in front of a detector. When the extremely powerful synchrotron beam hits the crystal, the protein inside diffracts the beam, resulting in a unique pattern of x-rays hitting the detector. The detector records the intensity of the beam and the pattern it forms, but not the beam's phase; all three are necessary to solve the protein's structure.

Researchers can calculate the phase information by approximating from a similar molecule that has already been solved.

"When you try to solve a new structure, in many cases part of the structure would be in some way similar to another molecule that had been previously determined," said [Piotr Sliz](#), the principal investigator for the [Structural Biology Grid](#).

Researchers feed the structure of a similar known protein, and the detector data from the unknown protein, into a computer program. If the known protein is close enough in structure to the unknown, the program can solve the structure of the unknown protein using a process called molecular replacement.

Normally, researchers know that their unknown protein is similar to a protein that has been solved and saved in the [Protein Data Bank](#). Sometimes, however, researchers know very little about their protein, or the subset of known proteins they've tried were not sufficiently similar. When that happens, they must turn to entirely different methods that would take weeks or even months of costly human time in the lab.

Invoking the grid

By using computational resources to compare unknown proteins to thousands of known proteins in only a few hours' time, [Ian Stokes-Rees](#) hoped to help structural biologists solve the structure of intractable proteins. Before they could convince biologists to use their application, however, they needed to provide proof of concept. They tried their process out by attempting to solve proteins that had been solved in 2008 using methods other than molecular replacement.

"About a year ago we stepped back and looked at the results that we'd collected from doing this process about a thousand times," Stokes-Rees said. "The evidence we had one year ago was not that compelling."

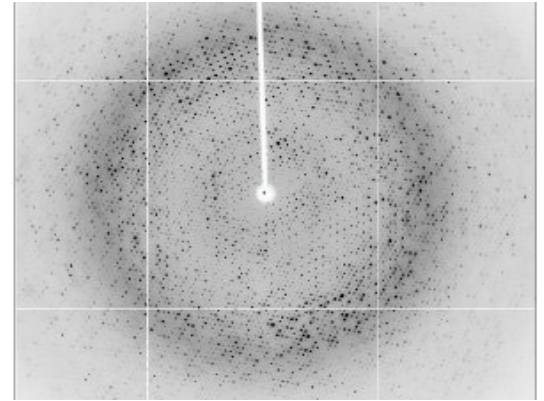
They decided to make two major changes to their method. First, they started using a different program called [Phaser](#) for the molecular replacement. Although Phaser is about ten times slower than the program they had been using, they hoped that it would produce better results. Second, they resolved to compare unknown proteins to the entire Protein Data Bank, rather than a large subset of the data bank.

These changes increased their computational needs from on the order of 100 hours per protein to about 20 000 hours, or from 4000 computations to 100000 per protein. Each run generates between 10 and 20 GB of data.

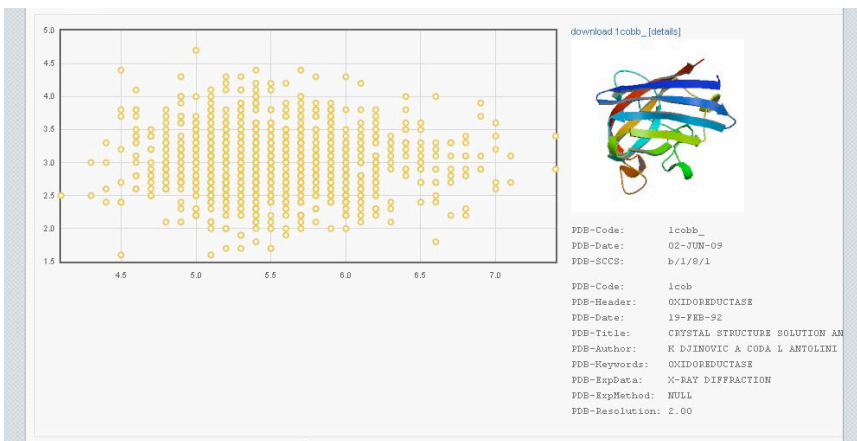
"We had to work pretty closely with the [Open Science Grid](#) people to determine how we could best manage the data, how we could manage the number of jobs, and how we could get enough computing time to complete these in less than a week," Stokes-Rees said.

They created a wrapper for Phaser, which is written in Fortran, and hooked it up to [DAGman](#), which served as a workflow management system. With this setup, they could send jobs to OSG. But they were still running into difficulties, as jobs would take nearly as long to start up as they took to run.

"Before we implemented [glideinWMS](#) we were really struggling to schedule 1000 jobs," Stokes-Rees said. "But once we were using glideinwms we were able to get up above ... 6000 jobs running."



Diffraction pattern obtained from crystals generated in [Sliz Laboratory](#) at Harvard Medical School. Image courtesy Sliz Laboratory, Harvard Medical School.



This screenshot shows results from an ongoing WMSR job on the Structural Biology Grid portal. Each orange circle represents a known protein that has been used as a possible template for the unknown protein. Users can click on any orange circle to see the protein displayed in the right-hand window. Screenshot by Miriam Boon.

The new technique was a success.

"We were able to show that there were certain cases that occur sufficiently frequently when this technique can be really valuable," Stokes-Rees explained. "So far we've found that about a quarter of the cases that we try, keeping in mind that the cases we try are generally people who are stuck, who have tried the conventional methods and are unable to get a good example for their structure . . . In about a quarter of those cases, we're able to find what look like strong candidate models."

This process will not, of course, work for proteins that are unlike anything in the Protein Data Bank. But for those cases where traditional methods have failed, this new process could save months or even years of work.

The [Wide Search Molecular Replacement](#) application is now available to users of the Structural Biology Grid; note that all job requests are screened to confirm that they merit significant computing time from OSG.

For more information about WMSR, please read the paper, which ran in the 22 November issue of the Proceedings of the National Academy of Sciences under the title, "[Protein structure determination by exhaustive search of the Protein Data Bank derived databases.](#)"

—Miriam Boon, iSGTW
www.isgtw.org



NSF INTERNSHIP AT HARVARD

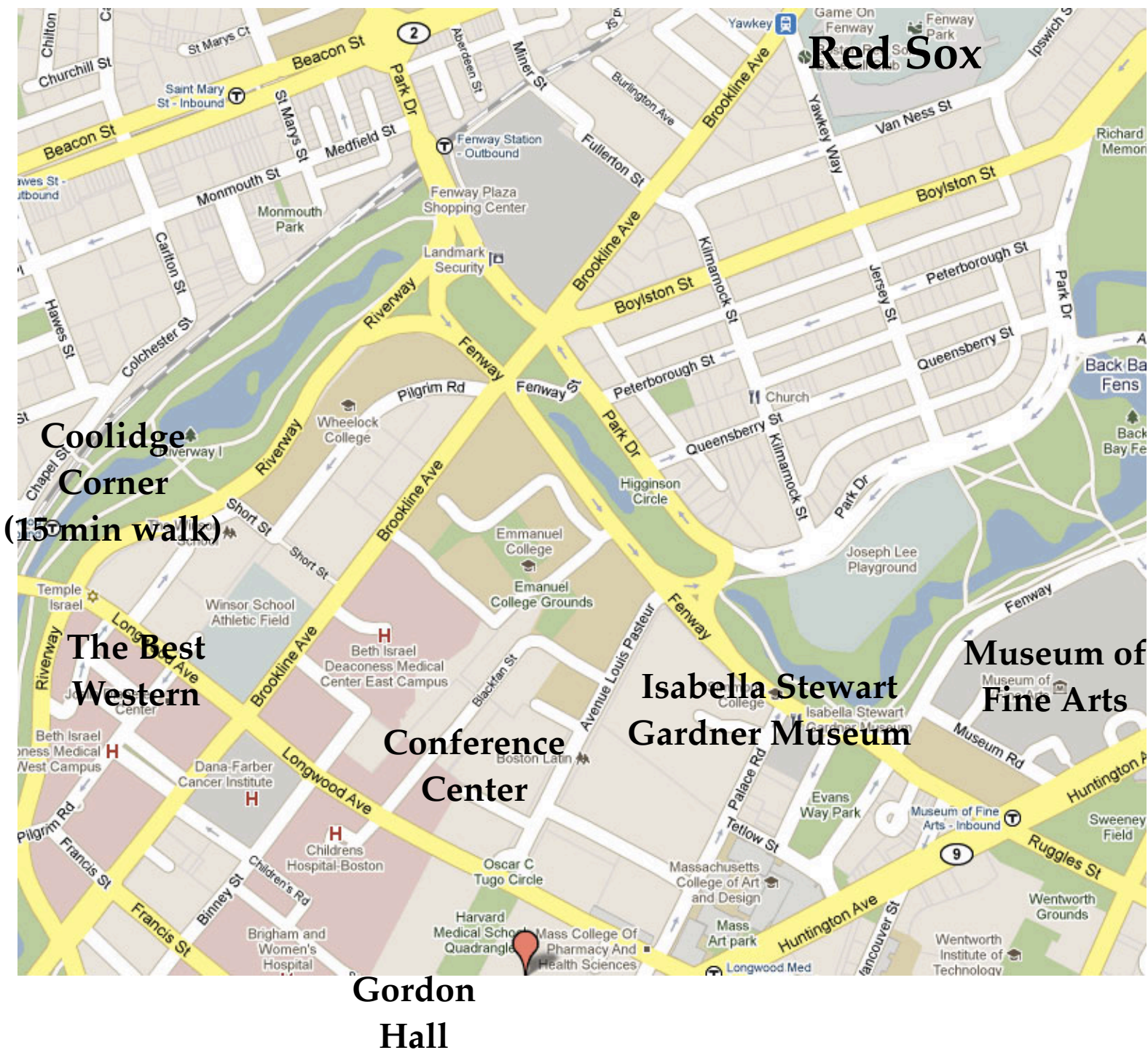
SUMMER POSITION FOR SCIENTIFIC SOFTWARE DEVELOPER

Based at **Harvard Medical School**, SBGrid (sbgrid.org) is a consortium of over 100 structural biology labs doing basic research into the structure and function of proteins. Over the last few years we have developed an infrastructure to manage very large computational and data intensive problems and are now expanding the web based interfaces to this infrastructure (portal.nebiogrid.org). The intern will work with an existing team of computational scientists to improve the existing tools and interfaces, focusing primarily on the front-end functionality and visualization. Key skills will be a knowledge of JavaScript, HTML and Python. It would be possible to continue working on this project part-time during the rest of the year. Through this internship you will contribute to new computational methods for protein structure studies and develop your skills in scientific software development.

Applicants aren't required to submit a formal CV. An email with a description of your experience, interest in the project, and (if available) some example web design work would be sufficient. Please e-mail your application to: apply@hkl.hms.harvard.edu

Deadline: April 4th 2011, 5:00 PM





Important Numbers:

- Best Western Inn at Longwood, 342 Longwood Ave, Boston, MA. Tel. 617-731-4700
- The Conference Center at Harvard Medical School, 77 Louis Pasteur Avenue, Boston, MA. Tel. 617-432-8993.
- Harvard Police Department: 617-432-1212
- Conference Organizers, email: ahm2011@sbgrid.org. Emergencies: 617-299-1455.