

Data Acquisition and Grid Computing For Africa

Dr. Julia Gray

on behalf of DOSAR and OSG

Outline

- History & OSG connection of DOSAR organization
- Grid computing outreach in Brazil
- DOSAR involvement in Africa
 - South Africa
 - SA Grid
 - University of Johannesburg and ATLAS
 - Paleontology project
 - ASP workshops
 - Ghana and updates
- Senegal and Future of Grid computing

What is DOSAR?

➔ **D**istributed **O**rganization for **S**cientific & **A**cademic **R**esearch

- Community and campus based Grid organization
- Its primary goal is spreading/enabling Grid use
 - DOSAR in Korean is the *God of Martial Arts*
- DOSAR evolved from the Fermilab National Accelerator Laboratory experiment DØ Remote Analysis effort
 - Coordinated group efforts in DØ simulation, reconstruction, and production

DOSAR History

- Ancient Times: DØ Remote Computing Era
 - SAM, DØ data management system: pre-2001
 - DØ remote analysis model proposed: Nov. 2001
 - Proposal for Regional Analysis Center (RAC) accepted and endorsed by DØ: Aug. 2002
 - UTA awarded MRI for RAC: June 2002
 - Formation of **DØ Southern Analysis Region (DOSAR 1.0)**: April 2003
 - DOSAR DØ Monte Carlo (MC) production begins
 - Activation of first RAC at UTA: Nov. 2003
 - Formation and Activation of DOSAR Grid for MC: April 2004

DOSAR History

- Modern Times: Beyond DØ
 - Transition to **D**istributed **O**rganization for **S**cientific & **A**cademic **R**esearch (DOSAR 2.0): April 2005
 - Active engagement with LHC experiments began
 - DOSAR VOMS installed at UTA: May 2005
 - DOSAR registered as a VO in OSG: July 2005
 - ATLAS distributed production and analysis system, Panda, implemented at OU and UTA: Jan. 2006
 - All groups engaged in LHC experiments
 - Outreach in Brazil and South Africa: 2008
 - DOSAR involvement with ASP: July 2011

Who is DOSAR?

- First Generation IAC's
 - University of Texas, Arlington
 - Louisiana Tech University
 - Langston University
 - University of Oklahoma
 - Tata Institute, India

- Second Generation IAC's
 - Cinvestav, Mexico
 - SPRACE – São Paulo Regional Analysis Center, Universidade Estadual Paulista, Brazil
 - University of Kansas
 - Kansas State University

- Third Generation IAC's
 - Ole Miss, University of Mississippi
 - Iowa State University
 - Louisiana State University
 - Oklahoma State University

- Fourth Generation IAC's
 - University of South Alabama
 - University of Johannesburg
 - GridUNESP

DOSAR Goals

- DOSAR seeks to bring together institutions which can/must use Grid technology to solve their computing problems (esp smaller institutions)
 - Computing needs increasing faster than funding for any one institution so harness strength in collaboration
 - DOSAR member institutions should serve as nucleation points for regional/statewide Grids
 - DOSAR should serve as a representative/liason for these institutions to the greater Grid community
 - DOSAR should help enable maximal use of available resources at the member institutions

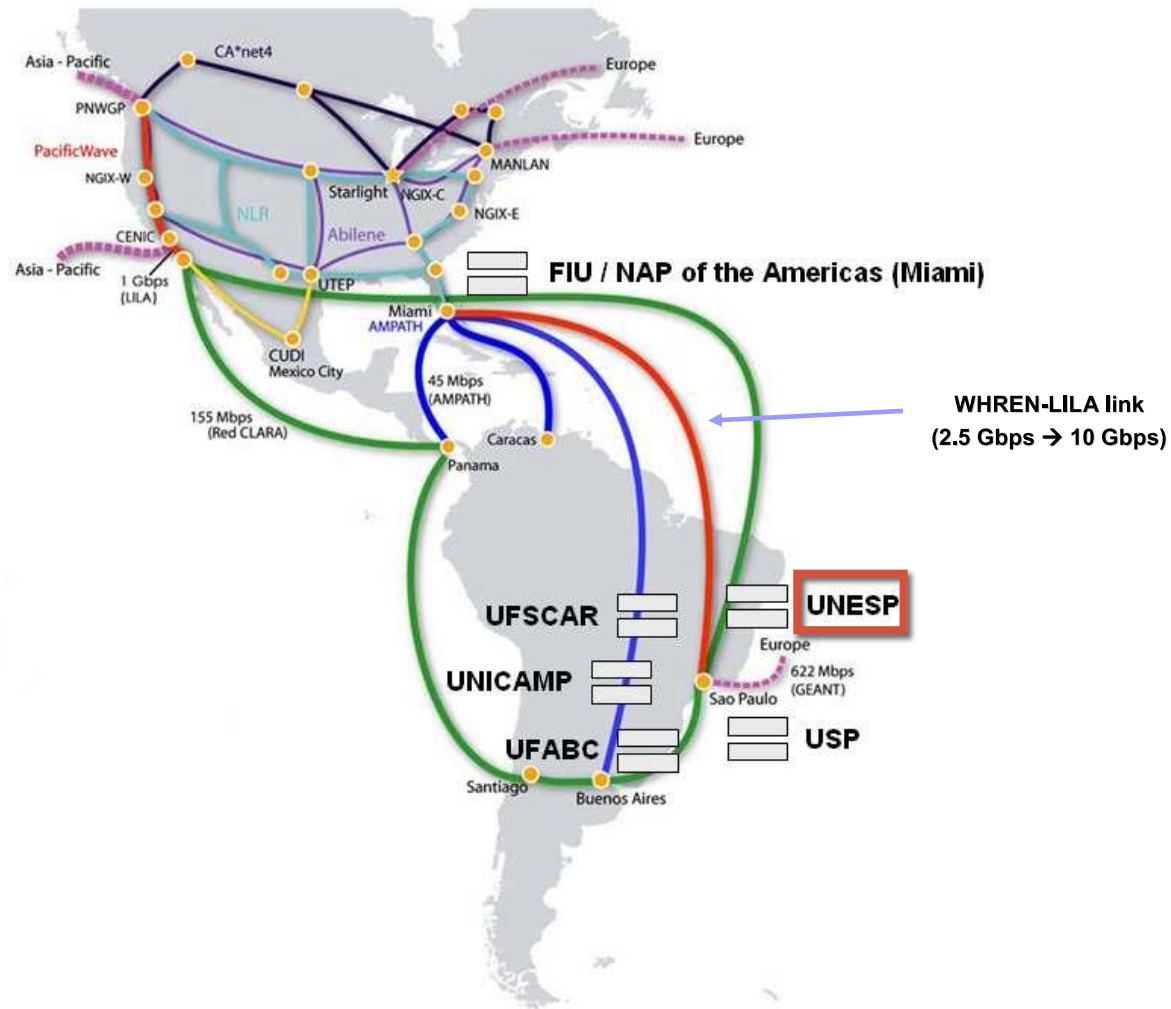
DOSAR and OSG

- DOSAR currently holds a seat on the Open Science Grid (OSG) counsel
- Engagement of non-HEP science
 - MPI jobs being submitted to NERSC by OU
 - Paleo project in South Africa.
- DOSAR works as the OSG Outreach and Educations organization
 - Grid computing workshops in South America and Africa
 - South Africa Grid School July 2008
 - São Paulo Grid School Dec 2008
 - GridUNESP Liasion to OSG: Fall of 2009
 - UJ/Wits strongly involved in Grid Education
 - ASP 2012 in Ghana

Outreach in Brazil

- DOSAR supported development of UNESP site and LHC CMS-Tier 2 computing facility
 - CMS Tier 2 and UNESP Grid now flourishing!
- Held several workshops to build local capacity
 - DOSAR Workshop, University of São Paulo: Sept. 2005
 - Brazilian LHC Computing Workshop I, University of São Paulo: May 2006
 - Brazilian LHC Computing Workshop II, University of São Paulo: Dec 2008
 - DOSAR Workshop, University of São Paulo: Sept. 2009
 - GridUNESP Workshop, University of São Paulo: Dec. 2009
 - OSG Workshop, University of São Paulo: Dec. 2010

Brazil Network



Brazil Grid Testbed

- Hardware: Sponsorship from leading manufacturers & suppliers
 - Intel (Xeon quad-core processors, network adapters, software tools)
 - SGI (Altix XE-310 barebones w/2 half-sized mainboards)
 - Kingston (DDR2 FB-DIMM memories)
 - Seagate (SATA-II hard disks)
- Software: Open-source tools and techniques
 - Virtualization provides server consolidation, minimizing hardware, physical space, and operational costs
 - Grid middleware based on the VDT (Virtual Data Toolkit)
 - VDT is a product of the OSG, a consortium of US universities, national labs, and computing centers

Outreach in South Africa

- DOSAR supported development of LHC computing facility and use of OSG software
 - Support for SA-ATLAS Grid facility
 - Non-LHC computing projects
- Held workshops to build local expertise
 - The South African Grid School, University of Witwatersrand: July 2008
 - DOSAR Workshop, University of Johannesburg: April 2010
- South Africa now active in outreach and training in Grid computing

SA Grid

- SA Grid started as the initiative of High Energy Physics group at UJ with support from DOSAR
 - User groups now include High Energy Physics, Astrophysics, Molecular Dynamics (Chemistry), Quantum Chemistry, Applied Mathematics, Numerical Studies (Engineering)
- National Grid to provide a national computing infrastructure to support scientific computing and collaboration
- Managed by a consortium of universities, national laboratories, and the Meraka Institute
 - Uses gLite middleware

UJ ATLAS Computing

- Jeremy Dodd (Nevis Labs) prepares proposal to OSG for HPC facility at University of Witwatersrand: Jan 2006
- Martin Cook, Claire Gray, Norman Ives attend Nebraska OSG school: Aug 2006
- OSG funding for HPC facility at Wits approved: Sept. 2006
- Grid site at Wits receive jobs and submit to OSGEDU VOs: Oct. 2007
- Developed model for SA participation in ATLAS via affiliation to Brookhaven: Oct. 2007
- Bruce Becker SA Grid development with gLite. SA-ATLAS agree to interoperability and resource sharing: 2008
- Horst Severini, Ben Clifford give OSG Grid School at Wits: July 2008
- Grid site rebuilt at UJ with OSG and gLite interoperability: Dec. 2008
- Wits joins UJ in SA-ATLAS: Jan 2010
- Currently contributing to HEP and computing training with OSG and ATLAS resources, gLite resources

South African Paleo Project

- Scan fossils at synchrotrons in X-ray radio-tomography and x-ray phase contrast tomography
 - Possible to acquire high resolution 3- D images and 3-D quantitative histograms of electron density.
 - Resolution can be submicron, and the specimen size can be as large as a skull
 - Raw sinogram can be a data file of the order of 1 TB.
 - To analyze the 3-D histogram to define and virtually extract objects representing all anatomical and geological features, with due consideration to the body of knowledge in the several fields of paleontology, anatomy, chemistry, geology, biology, applied mathematics, physics and computer science is the bottleneck step of several years.
 - Developing algorithms to facilitate this analysis by unifying the knowledge and skill from these fields is an important enabling step to properly beneficiate this new
- The SA Paleo Project to develop the mathematics for the algorithms described above
 - Facilitate a higher degree of automation of the analysis of paleo-tomography data with a higher degree of fidelity.
 - The new algorithms would make full use of the quantitative nature of the data, the physics that it represents, the anatomical context, the diagenesis process and other geological, chemical and biologically relevant aspects.
 - Deliver a toolkit for processing based on modern computing practice.
- The project is very impactful for international paleontology
- Significant for Southern African Heritage studies

DOSAR & African School of Physics

- In 2012, dedicated 3 day workshop added to the end ASP
 - Distributed and high-throughput computing, security, processing, and distributed storage covered during school
 - Partnership to install Grid software on existing cluster
 - Computing cluster donated to KNUST in Kumasi, Ghana after ASP
 - Ghana is looking to join ATLAS and CMS at CERN
- Now African Grid School integrated part of ASP curriculum
 - Two days of Grid lectures as part of the school in Senegal



Senegal the Future of Grid Computing

- The future of Grid computing is Graphical Processing Unit (GPU)
 - Can process images, simulations, and data quickly
- Senegal is receiving a donation of computers from CERN
 - Size is sufficient for a Tier 3 site for LHC experiments
- Distributed computing and storage useful for digital library
 - Over 18,000 documents currently stored at UCAD
 - Library access to Senegalese university network and consortium of 8 countries
 - Grid computing can enable the storage, search, and access documents as one system no matter where they are stored or accessed