UAB IT
Research Computing Update

UAB IT Research Computing Day

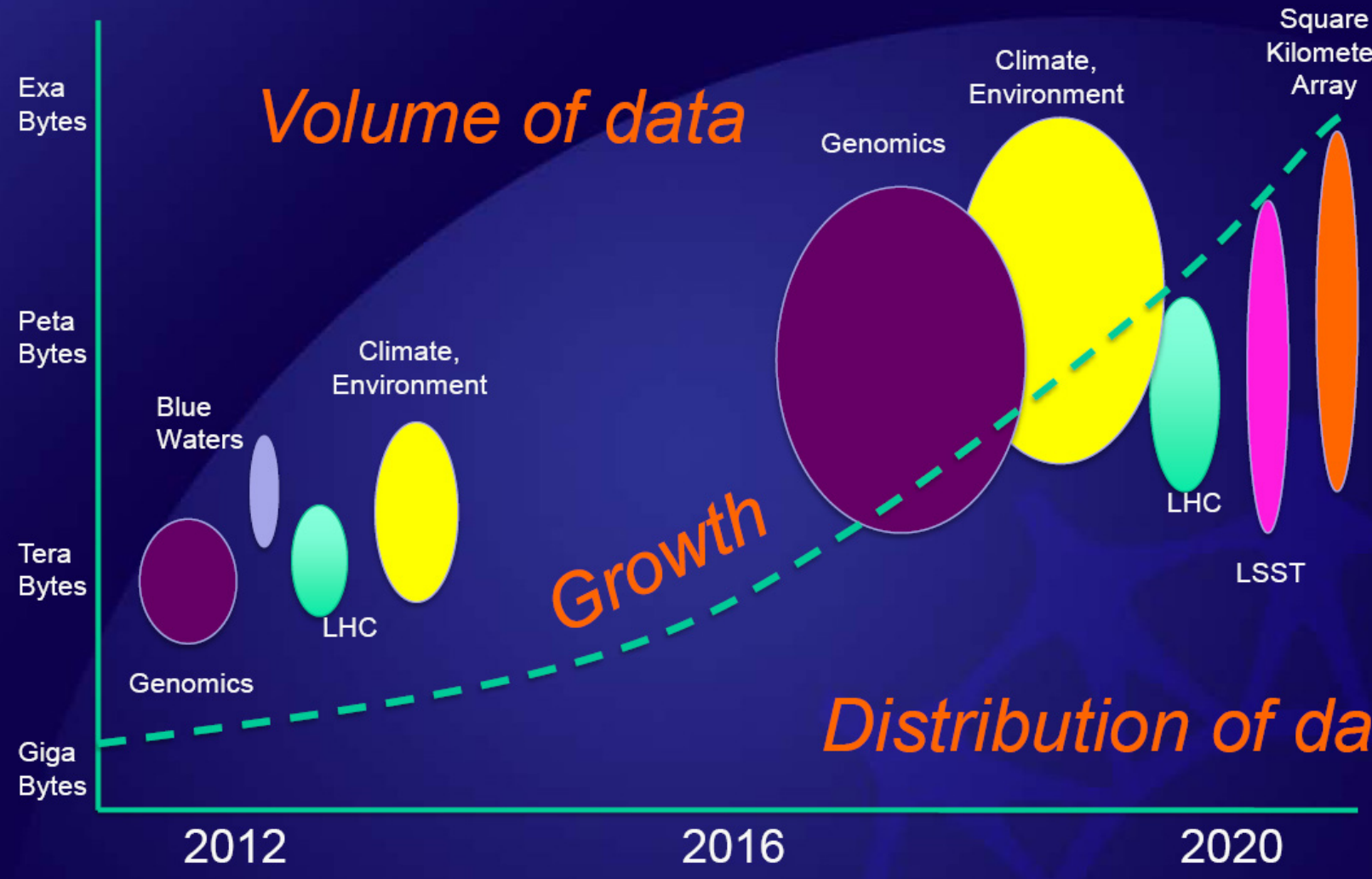
September 15,2011

UAB INFORMATION TECHNOLOGY

Discussion Topics

- Research Applications
- Developer Tools
- Hardware
- Networks
- People
- Current Projects

Data Challenges



Interoperability of Data

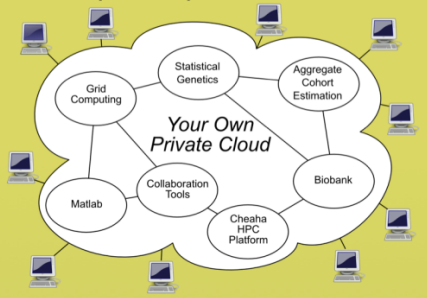
"NSF CI Framework for 21st Century Science & Engineering (CIR21)", Alan Blatecky, http://www.nsf.gov/attachments/119046/public/CIF21_Presentation.pdf

Research Computing Support

*brought to you by **UAB IT***

UAB Research Cloud

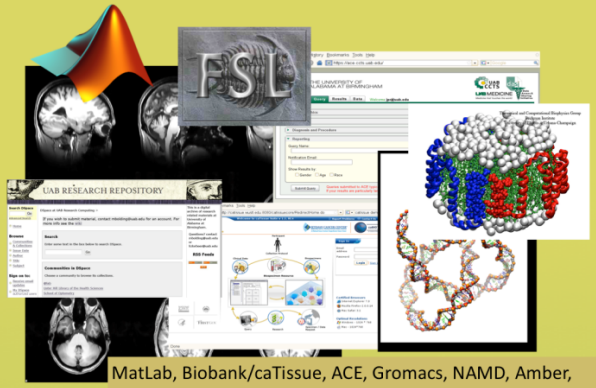
platforms , computing and storage to power your research



The research cloud is designed to support private workspaces with access to the tools you need where ever you need them. Take your lab anywhere, virtually.

Research Apps

solutions built to order



MatLab, Biobank/caTissue, ACE, Gromacs, NAMD, Amber, FSL, GNU Scientific Toolkit, and much more

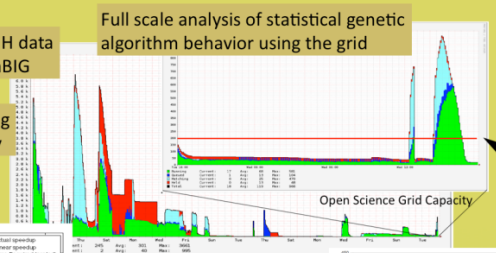
A large variety of High Performance Computing apps help make sense of piles of data and model complex systems. Designer apps help speed up tasks and fill in gaps whether its world wide computing, tissue banking, cohort discovery, or data archiving.

Results Delivered

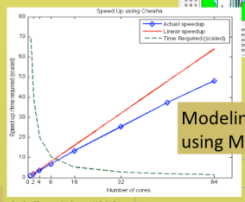
Connections to NIH data networks using caBIG

Full scale analysis of statistical genetic algorithm behavior using the grid

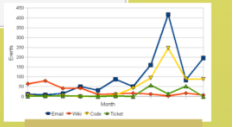
On-line tissue banking and Cohort discovery



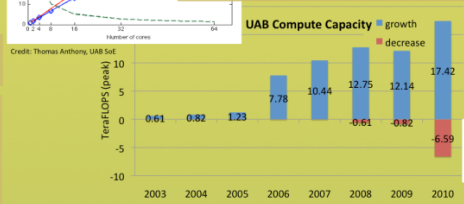
Open Science Grid Capacity



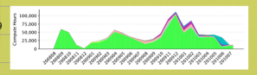
Modeling sensor networks 48x faster using Matlab cluster computing



A network of experts across organizations



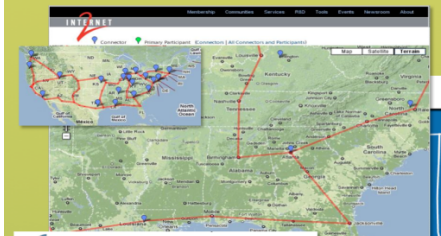
Cheaha: Increased Cluster Utilization



A growing resource fabric: combining funds from research Awards and sustained UAB IT investments to fund on-going expansion of compute, network, and storage capacity and maximize the impact and availability of acquisitions.

UASRON

your connection to the world



UASRON is a very high bandwidth lambda network that connects the UA System to National Lambda Rail and Internet2. UAS RON is a collaboration between the UA System, NASA, and Southern Light Rail (SLR)@GaTech.

University of Alabama System Owned and Operated, UASRON is a Dense Wave Division Multiplex (DWDM) Network, analogous to a 40-Lane, bi-directional, Data Highway connecting UAB, UAH, UA, ASA, to Atlanta and Nashville. Current technology allows each lane (channel) to carry a payload of 10Gigabits-per-second.

Developer Accounts

an open platform to build research

MediaWiki, WordPress, Mailing List, Trac (docs & bugs), Git(file versioning)

Engage your peers. Our platforms, languages, and collaboration tools make it easy to build solutions, document code, record changes, track bugs and collaborate across organizations. Mediawiki, WordPress, mailing lists, Trac, Git, your favorite languages, and slew of other tools: we live on the same platform and run our lab that way too.

The Team

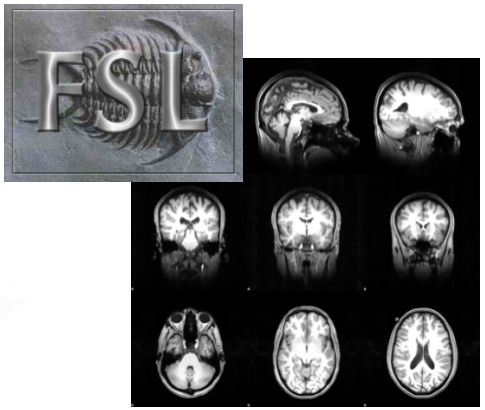
An engaged professional staff collaborate with investigators to maximize the impact of research

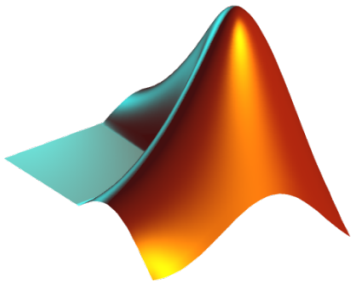
Bob Cloud	David Shealy	PuriBangalore	John-Paul Robinson
Executive Director Infrastructure Services UAB IT	Director Research Computing Physics / UAB IT	Consultant to Research Computing CIS / UAB IT	System Architect Research Computing UAB IT

Mike Hanby	PoornimaPochana	ShantanuPavgi	Bill Bradley
Information Systems Specialist Engineering / UAB IT	Programmer/Analyst Research Computing UAB IT	Programmer/Analyst Research Computing UAB IT	System Programmer Infrastructure Services UAB IT

Research Applications
solutions build to order

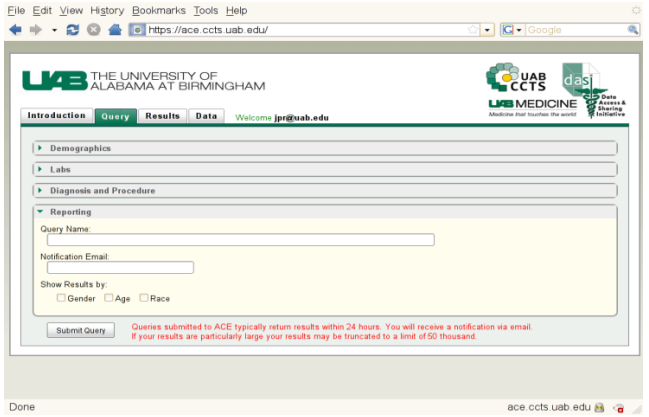
Neuroimaging





Data Analysis with MATLAB

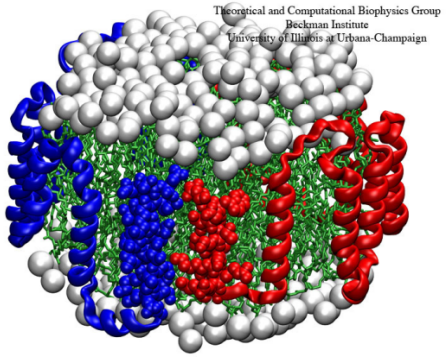
Patient Cohort: ACE



Digital Archives

NextGen Sequencing

Molecular Dynamics

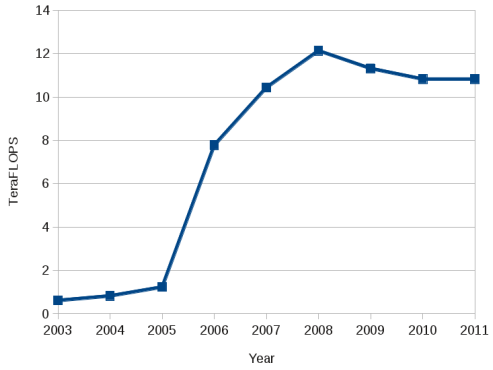


Fast apps help make sense of piles of data and model complex systems. Designer apps improve efficiency and fill the gaps whether its nextgen sequencing, cohort discovery, data archiving, or computing across the world.

Results Delivered

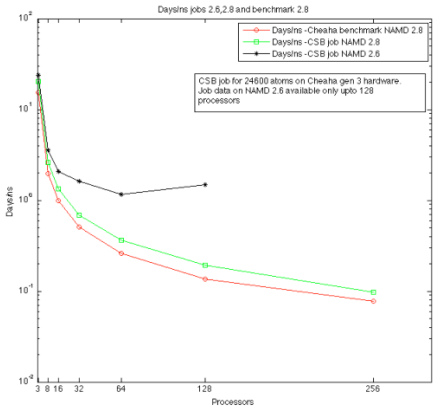
HPC TeraFLOPS

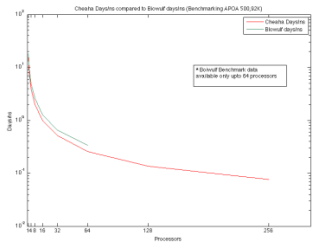
2003-2011



Application Performance

8x Increase for NAMD





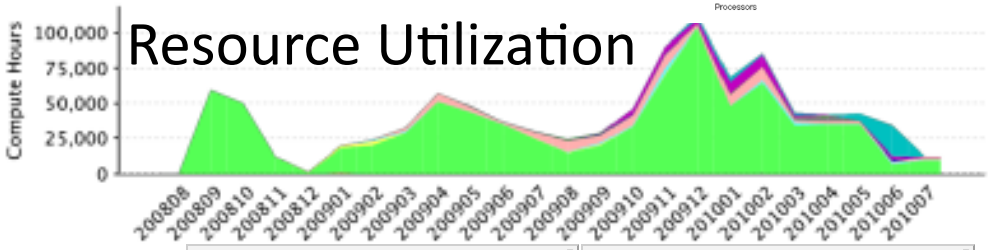
NAMD Comparison

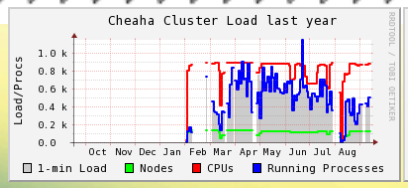
System	Atoms	ns/day	CPU count
UAB BlueGene	246,000	0.80	256
Cheaha with SIG	235,000	0.88	32

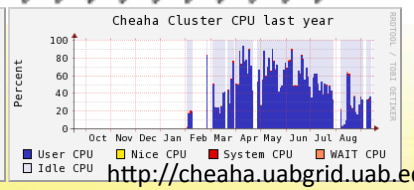
Compute Community

Galaxy and MATLAB

Resource Utilization

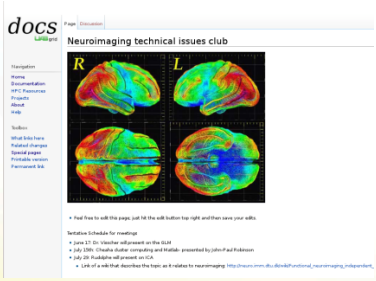






http://cheaha.uabgrid.uab.edu/ganglia

Collaboration Support



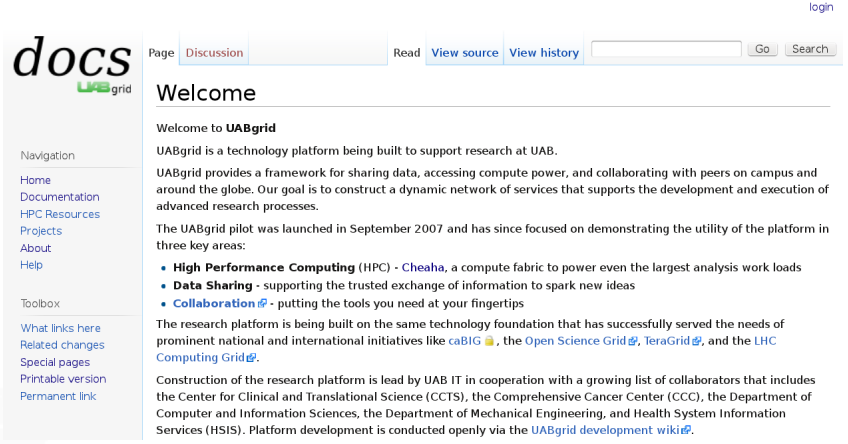
Wiki & Project Spaces:

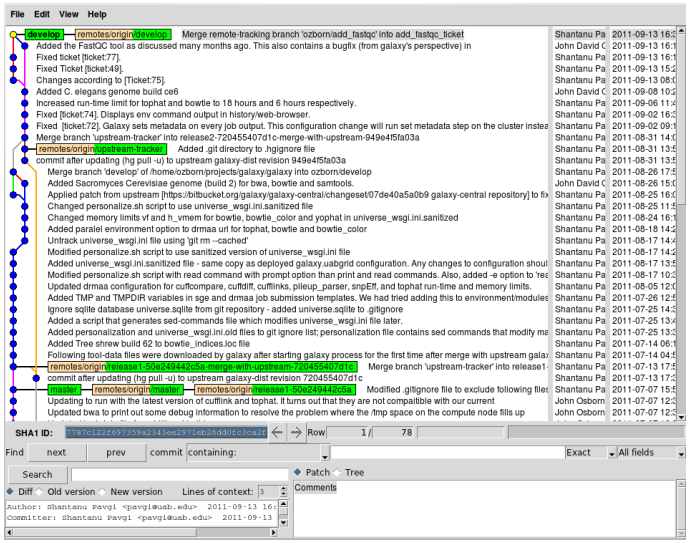
- Neuroimaging Club
- MATLAB docs
- Digital Archives
- NLP
- Cerner PowerInsight

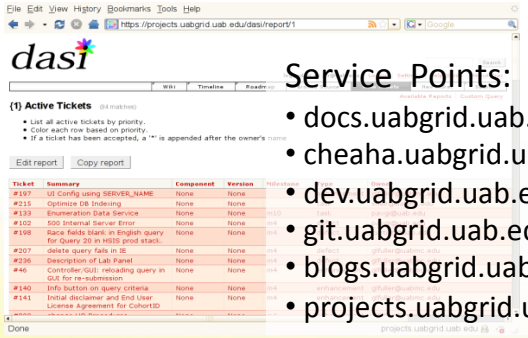
A growing resource fabric combines dollars from research Awards and UAB IT investments to fund on-going expansion of compute, network, and storage capacity and maximize the impact and availability of acquisitions.

Developer Tools

an open platform to build research





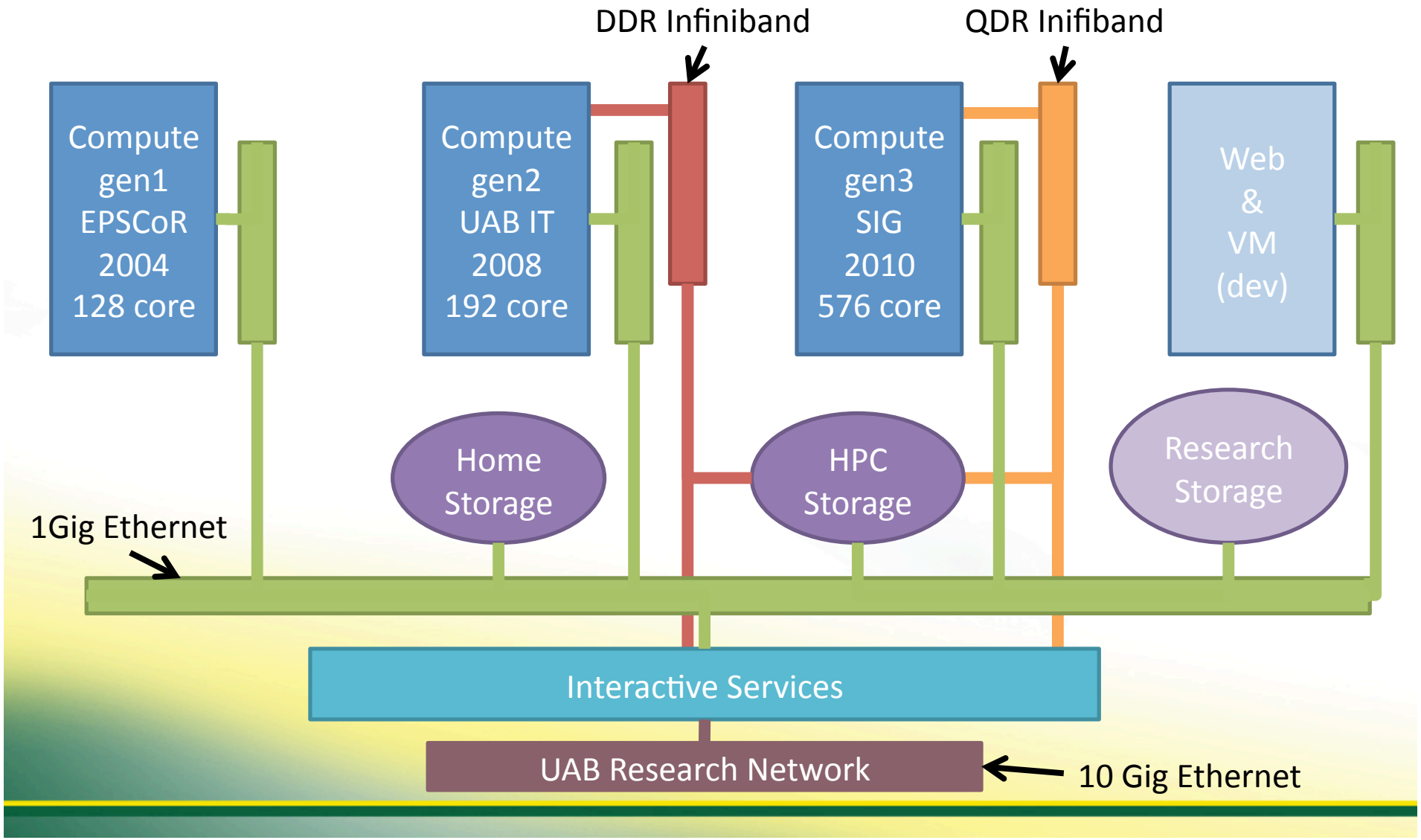


Service Points:

- docs.uabgrid.uab.edu
- cheaha.uabgrid.uab.edu
- dev.uabgrid.uab.edu
- git.uabgrid.uab.edu
- blogs.uabgrid.uab.edu
- projects.uabgrid.uab.edu/<pname>

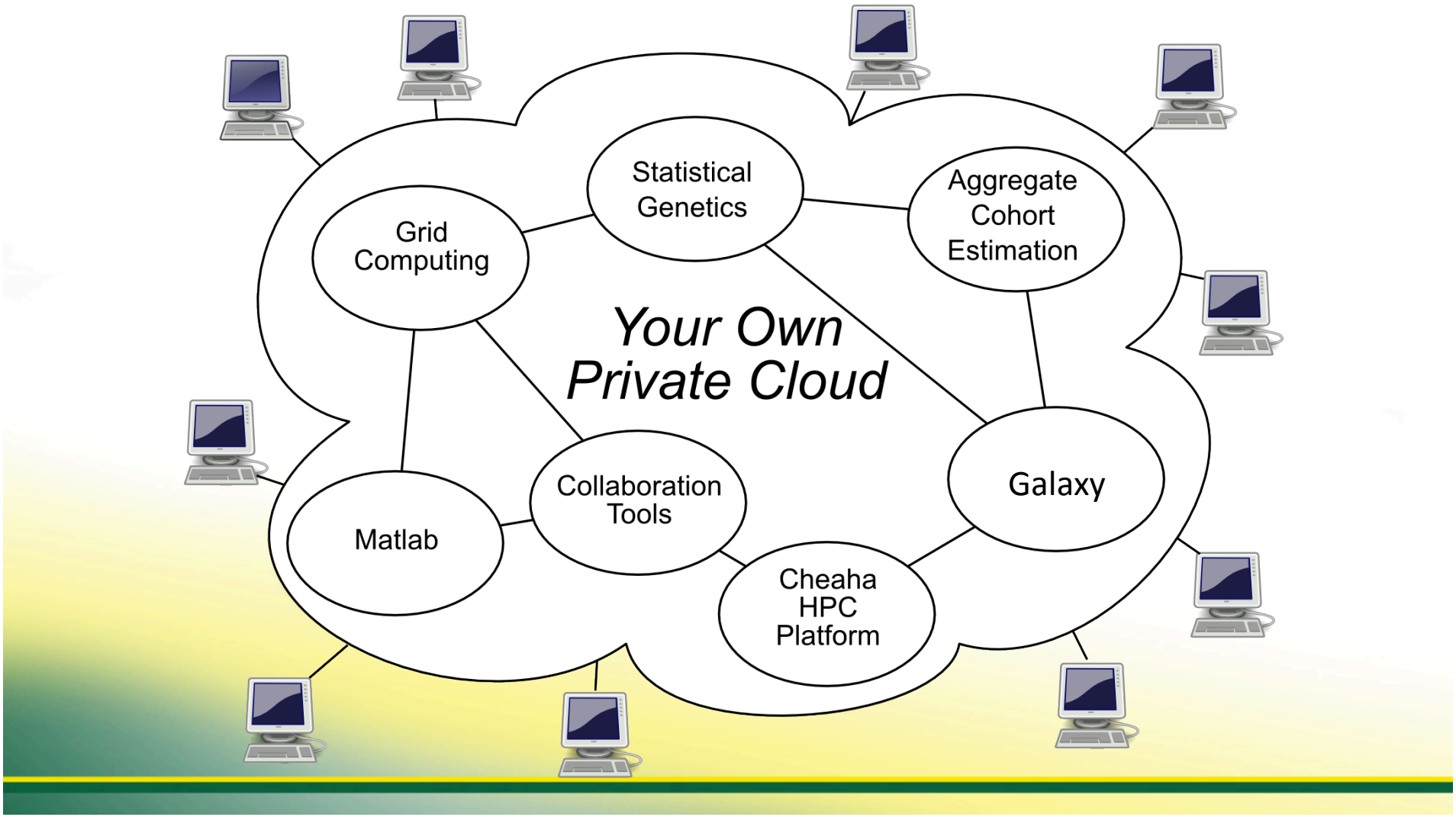
Engage your peers. Our platforms, languages, and collaboration tools make it easy to build solutions, document code, record changes, track bugs and collaborate across organizations. Mediawiki, WordPress, mailing lists, Trac, Git, your favorite languages, and slew of other tools: we live on the same platform and run our lab that way too.

Research Hardware Platform



UAB Research Cloud

platforms, computing and storage to power your research



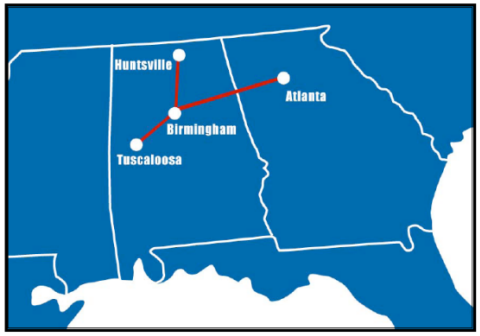
UAB Research Network and UASRON

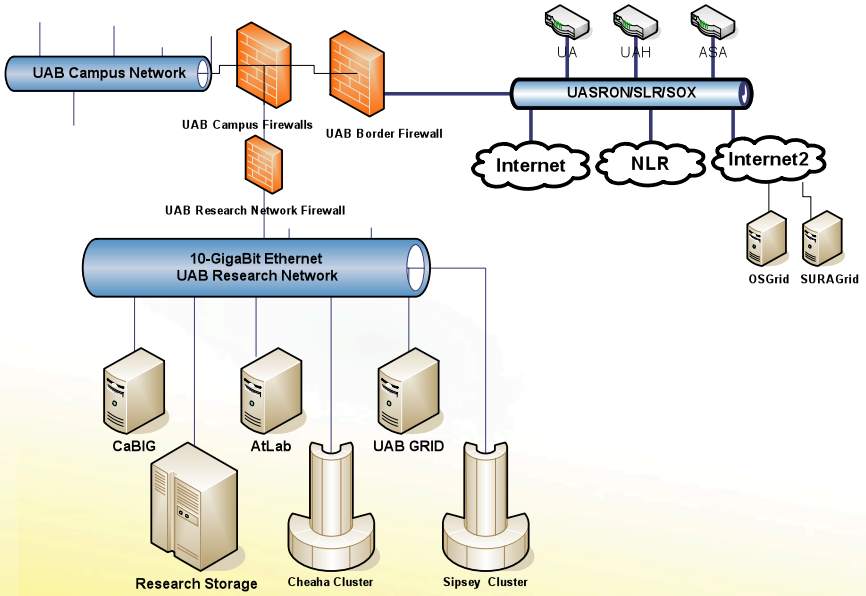
connections to the world

Chan Soon-Shiong Institute for Advanced Health announces **NATIONAL HEALTH INTERNET** – $100 million funding of National LambdaRail to bring the power of supercomputing and genomic analysis to the point of care

UASRON: Leased Dark Fiber

A very high bandwidth lambda network that connects the UA System to National Lambda Rail and Internet2. UAS RON is a collaboration between the UA System, NASA, and Southern Light Rail (SLR)@GaTech.





Internet2: Consortium of over 250 universities, research labs, government, and corporations. Promotes collaboration and education. Worldwide partnership, not limited to US/North America. More than 50-International Partners.

University of Alabama System

Regional Optical Network (RON)

- UASRON is a Privately Owned ,10-Gigabit-Ethernet, Dense Wavelength Division Multiplex (DWDM) network.
- Utilizes Leased Dark Fiber from: Level3, EarthLink, Charter Communications, and XO in Nashville.
- Connects UA System to the National LambdaRail (NLR) and Internet2 in both Atlanta and Nashville.
- Owned and Operated by the UA System.
- Utilized by UA, UAB, UAH ,NASA, Alabama Super Computer Authority, and Southern Light Rail (SLR) @GaTech.

New NLR Partner for FY2012

- INSTITUTE FOR ADVANCED HEALTH ANNOUNCES HIGH PERFORMANCE SECURE "NATIONAL HEALTH INTRANET"

- Chan Soon-Shiong Institute for Advanced Health Provides $100 million Funding of National LambdaRail to Bring the Power of Supercomputing and Genomic Analysis to the Point of Care

UASRON Description

- Dense Wave Division Multiplex Network,
- DWDM is Analogous to a 40-Lane, bi-directional, Data Highway connecting UAB, UAH, UA, ASA, to Atlanta and Nashville.
- Current technology allows each lane (channel) to carry a payload of 10Gigabits-per-second.

History

- Approved by UA Board of Trustees in 2005
- Completed initial Build-out in 2007.
- Funded by a Partnership between the UA System and NASA.
- Twenty-Year IRU Fiber Lease guarantees long term ownership , and un-restricted partnerships.
- Principal Vendors: Adva Optical, Southern Light Rail (SLR), Level3, Earthlink, Charter Communications, Telx.

Wavelengths (Lambdas) in Operation

- UAB Hub provides 10 Gigabit-Ethernet to 4Nodes:
  University of Alabama (Tuscaloosa), ASA (Huntsville) , and SLR/SOX(Atlanta). Sox (Nashville) was added in Spring-2011

  SOX (Ga. Tech.)provides access to: Internet2, National Lambda Rail (NLR), and Commodity Internet.
- Other UASRON Lambdas include:
  - 3ea ASA/AREN 10G Ethernet, connecting Huntsville to Birmingham, and Huntsville to SOX-Atlanta , and Level3-Nashville.
  - 2ea UA to SOX-Atlanta
  - 1ea UAH to SOX-Atlanta
  - 2ea NASA to NASA-Atlanta
  - 3ea ORNL Local Channels in Nashville

Future Wavelength Plans for UASRON

- SOX-Nashville: New SOX Node to be established in Nashville, Requires Local Channels for NLR, Internet2.
- UAB to SOX-Nashville for Redundancy.
- SOX-Nashville to SOX-Atlanta.

Internet2

- Started in 1996
- Consortium of over 250 Universities, Research Labs, Government, and Corporations.
- Also supports K-20 Education Community.
- Network and Community Services for
  - Advanced network research
  - Revolutionary Internet Technologies
  - Promotes Collaboration and Education

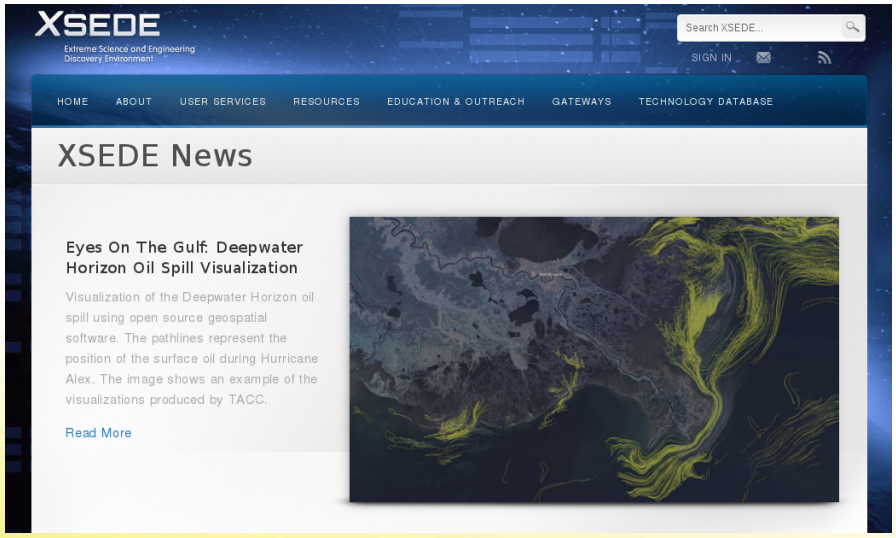
  --Worldwide Partnership, not limited to US/North America. More than 50-International Partners.

National Research Resources

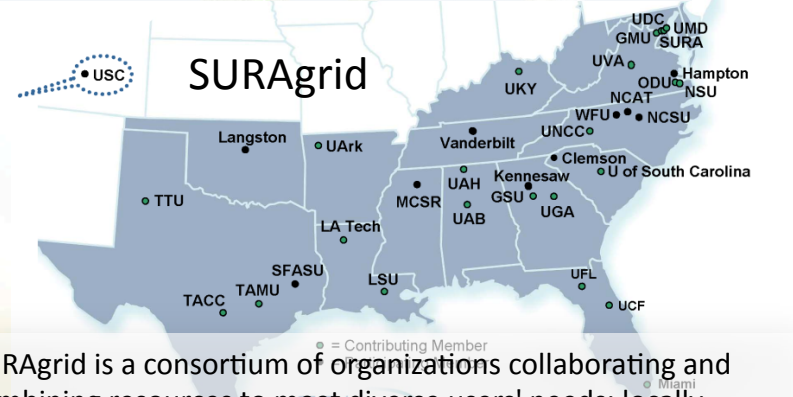
XSEDE, Open Science Grid, SURAgrid

The Open Science Grid (OSG) advances science through open distributed computing. The OSG is a multi-disciplinary partnership to federate local, regional, community and national cyberinfrastructures to meet the needs of research and academic communities at all scales.

Open Science Grid

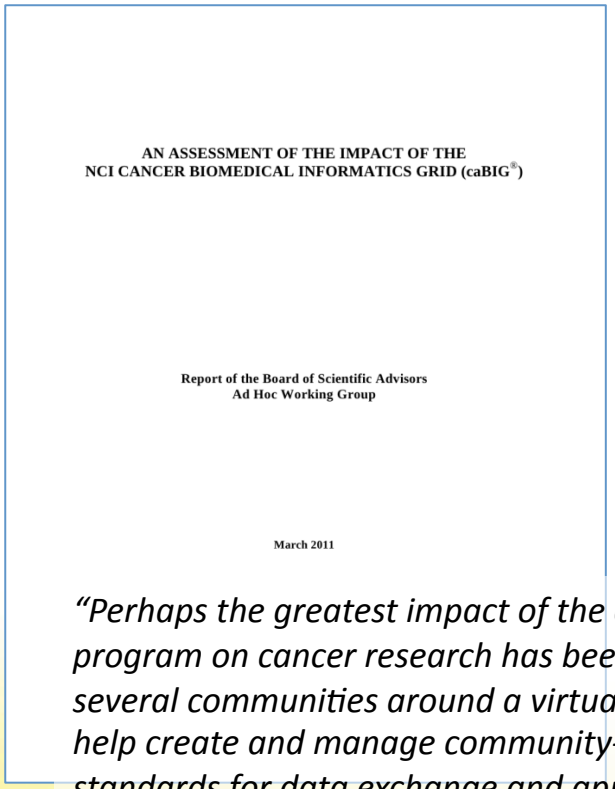


XSEDE is a five-year, $121-million project supported by the National Science Foundation. It replaces and expands on the NSF TeraGrid project.



SURAgrid is a consortium of organizations collaborating and combining resources to meet diverse users' needs: locally contributed resources, project-specific tools and environments, highly specialized or HPC access, and gateways to national and international cyberinfrastructure.

Collaboration Across UAB

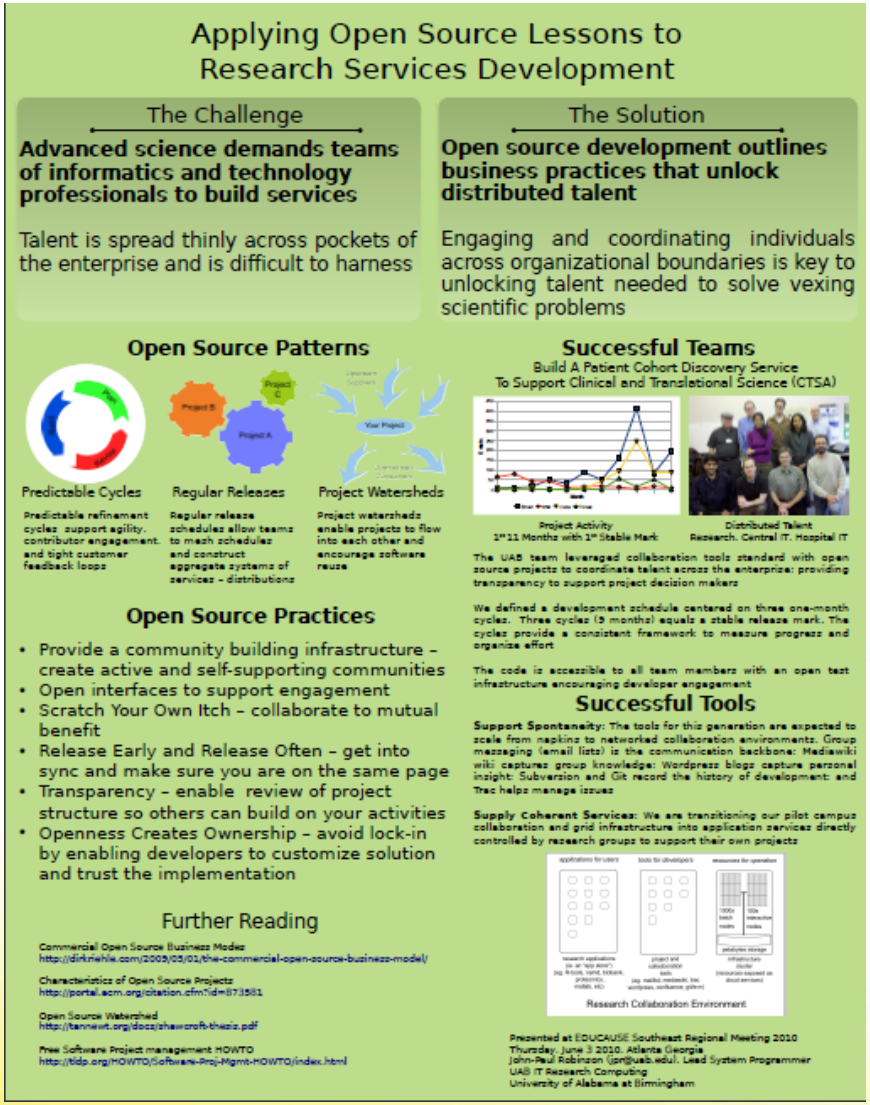


AN ASSESSMENT OF THE IMPACT OF THE
NCI CANCER BIOMEDICAL INFORMATICS GRID (caBIG®)

Report of the Board of Scientific Advisors
Ad Hoc Working Group

March 2011

"Perhaps the greatest impact of the caBIG program on cancer research has been to gather several communities around a virtual table to help create and manage community-driven standards for data exchange and application interoperability."



Applying Open Source Lessons to Research Services Development

The Challenge

Advanced science demands teams of informatics and technology professionals to build services

Talent is spread thinly across pockets of the enterprise and is difficult to harness

The Solution

Open source development outlines business practices that unlock distributed talent

Engaging and coordinating individuals across organizational boundaries is key to unlocking talent needed to solve vexing scientific problems

Open Source Patterns

Predictable Cycles

Predictable refinement cycles support agility, contributor engagement, and tight customer feedback loops

Regular Releases

Regular release schedules allow teams to mesh schedules and construct aggregate systems of services – distributions

Project Watersheds

Project watersheds enable projects to flow into each other and encourage software reuse

Open Source Practices

- Provide a community building infrastructure – create active and self-supporting communities
- Open interfaces to support engagement
- Scratch Your Own Itch – collaborate to mutual benefit
- Release Early and Release Often – get into sync and make sure you are on the same page
- Transparency – enable review of project structure so others can build on your activities
- Openness Creates Ownership – avoid lock-in by enabling developers to customize solution and trust the implementation

Further Reading

Commercial Open Source Business Model
http://dirkriehle.com/2009/03/01/the-commercial-open-source-business-model/

Characteristics of Open Source Projects
http://portal.acm.org/citation.cfm?id=873581

Open Source Watershed
http://tanniwt.org/docs/shawcroft-thesis.pdf

Free Software Project management HOWTO
http://tldp.org/HOWTO/Software-Proj-Mgmt-HOWTO/index.html

Successful Teams

Build A Patient Cohort Discovery Service
To Support Clinical and Translational Science (CTSA)

Project Activity
1st 11 Months with 1st Stable Mark

Distributed Talent
Research, Central IT, Hospital IT

The UAB team leveraged collaboration tools standard with open source projects to coordinate talent across the enterprise: providing transparency to support project decision makers

We defined a development schedule centered on three one-month cycles. Three cycles (3 months) equals a stable release mark. The cycles provide a consistent framework to measure progress and organize effort

The code is accessible to all team members with an open test infrastructure encouraging developer engagement

Successful Tools

Support Spontaneity: The tools for this generation are expected to scale from napkins to networked collaboration environments. Group messaging (email lists) is the communication backbone; Mediawiki wiki captures group knowledge; Wordpress blogs capture personal insight; Subversion and Git record the history of development; and Trac helps manage issues

Supply Coherent Services: We are transitioning our pilot campus collaboration and grid infrastructure into application services directly controlled by research groups to support their own projects

Research Collaboration Environment

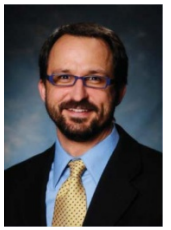
Presented at EDUCAUSE Southeast Regional Meeting 2010
Thursday, June 3 2010, Atlanta Georgia
John-Paul Robinson (jpr@uab.edu), Lead System Programmer
UAB IT Research Computing
University of Alabama at Birmingham

IT Research Computing Team

Bob Cloud
Executive Director
Infrastructure Services
UAB IT

David Shealy
Director
Research Computing
Physics / UAB IT

Puri Bangalore
HPC Consultant to
Research Computing
CIS / UAB IT



John-Paul Robinson
System Architect
Research Computing
UAB IT

Mike Hanby
Information Systems
Specialist
Engineering / UAB IT

Poornima Pochana
Programmer/Analyst
Research Computing
UAB IT

Shantanu Pavgi
Programmer/Analyst
Research Computing
UAB IT

Bill Bradley
System Programmer
Infrastructure Services
UAB IT

Engaged professional staff collaborate with investigators to maximize the research impact

Current Projects

- Documentation!
- Enhance platform features
- Grow MATLAB use and support
- Build out NGS tools and services
- Pilot Condor compute fabric
- Identify data management solutions

Proposed Plans
UAB Research Computing Data Storage

- Users: PI's of grants, users of HPC environment (Cheaha), and Group Access
- Target initial maximum allocation of 1TB
- Access Technologies: BlazerID authenticated; Network attached; WebDAV, Drop Box, & SFTP
- Data Retention: indefinite retention online; archive to tape or other media is available for a fee upon request.