

Sustainable Funding Models for Academic High Performance and Research Computing

...Or at least Random Musings About Them

Dan Stanzione

Director

10/4/06



IRA A. FULTON SCHOOL OF ENGINEERING
HIGH PERFORMANCE COMPUTING

Outline

- Why talk about this now?
- Existing Models and National Examples
 - Why people fund? us, and what's the return
- The ASU Experience: Lessons Learned Building a Sustainable Center
- Discussion: What's the Right Model?
- Where to go from here



IRA A. FULTON SCHOOL OF ENGINEERING
HIGH PERFORMANCE COMPUTING

The Givens

- For this Audience, let's skip the motivation speech, and make it a given that HPC / Supercomputing is critically important to continued research progress
 - Capacity: accelerating time to result
 - Capability: otherwise unachievable results
 - Replacing experimentation when too costly.
- We all think we should do big research computing, even if we differ on where and how.



IRA A. FULTON SCHOOL OF ENGINEERING
HIGH PERFORMANCE COMPUTING

Context

- Federal Funding in this area is NOT getting scarcer (a rarity), but is getting more concentrated...
- NSF Petascale, DOE NLCF moving to even fewer, more massive centers
 - NSF will make \$260M in new HPC awards in FY07...at two sites (or maybe one).
 - For reference, total CISE spending is <\$600M, and represents 81% of all academic CS research
- For most of us, these awards aren't helpful for developing local resources



IRA A. FULTON SCHOOL OF ENGINEERING
HIGH PERFORMANCE COMPUTING

Centers

- All the (NSF) Centers have great value, and all have made massive contributions to the field. The Teragrid serves many needs for big users *but*
- Data Migration is an issue, particularly if data is generated locally
- The environment isn't flexible for apps with, *ahem* "novel" requirements.
- Tough to do nimble allocations to seed research growth; same for education
- Teragrid delivers *cycles* to your users, but not the rest of what a computation center should, such as:
 - (Human) Services and Expertise
 - Visualization
 - Data life cycle management
- *Centers are a partial solution!*



IRA A. FULTON SCHOOL OF ENGINEERING
HIGH PERFORMANCE COMPUTING

What Others have done...

- **Final Report: A Workshop on Effective Approaches to Campus Research Computing Cyberinfrastructure**
- <http://middleware.internet2.edu/crcc/docs/internet2-crcc-report-200607.html>
 - Sponsored by Internet2 and NSF
- **Conclusions offered a list of problems, but no solutions, including:**
 - Major negative reinforcements exist in the current environment. For example, grant solicitations at several major funding agencies seem to favor "autonomous, small clusters in closets" over more sustainable and secure resources.
 - There is increasing conflict between research computing systems and campus security procedures.
 - The profile of researchers who use campus research computing cyberinfrastructure seems different from those who use national resources
 - There are several campuses which are very active in supporting local research cyberinfrastructures, but even they are challenged with chronically insufficient resources
 - *(Lack of) attention by funding agencies to this problem was pointed out 6 times in the executive summary - which tell you the goal of the meeting*



IRA A. FULTON SCHOOL OF ENGINEERING
HIGH PERFORMANCE COMPUTING

Where Money Comes From

- Center grants
- Private Funds
- State Funds
- Research Office
- Central IT
- Smaller Academic Units
- Other grants / individual faculty



IRA A. FULTON SCHOOL OF ENGINEERING
HIGH PERFORMANCE COMPUTING

HPC is important. So is world peace. Why should we get resources?

- HPC is in constant competition with other research/university priorities
- IT is already a huge line item in the budget of every university and every unit in it (ASU: \$90M, half distributed).
- It is true HPC is critical to research progress, but somebody will make the same argument about the electron microscope down the hall, or wetlab space, or, sequencers, or...
- And just because it's important, doesn't mean anyone cares...



IRA A. FULTON SCHOOL OF ENGINEERING
HIGH PERFORMANCE COMPUTING

What are the Funders After? (ROI, CYA, LMA, ED)

- The VP for Research - ROI
 - Myth: VPRs want you to make great discoveries
 - Reality: VPRs want you to discover you've been externally funded.
- The research office is not a funding agency; a good research office invests in centers that provide a Return-on-Investment. For HPC centers, this means either:
 - Win your own grants
 - Drive research wins (how do you measure impact?)
- Exemplar: TACC



IRA A. FULTON SCHOOL OF ENGINEERING
HIGH PERFORMANCE COMPUTING

What are the Funders After? (ROI, CYA, LMA, ED)

- Central IT: CYA
 - It is not in the best interest of university IT for each faculty member to run their own cluster and fileserver.
 - Security: IT is still responsible for the network
 - Data Integrity: Who will the Inspector General call when federally funded research data is lost?
 - Lots of support calls on unmanaged systems: This is a no-win situation, even if you refuse



IRA A. FULTON SCHOOL OF ENGINEERING
HIGH PERFORMANCE COMPUTING

What are the Funders After? (ROI, CYA, LMA, ED)

- State Legislatures: ED
 - Economic Development; states want to create jobs. They will care if you bring in companies or create new companies (employing people in your center does not count as job creation :)
- Academic Units: LMA
 - Leave Me Alone; Some deans will spend some money on this to quiet wailing faculty



IRA A. FULTON SCHOOL OF ENGINEERING
HIGH PERFORMANCE COMPUTING

A Few Myths

- F&A will pay for this!
 - Research Computing is distinctly NOT in the indirect cost rate... F is physical plant only, and is break even... A is everything else, and is usually a money loser. Plus, these dollars pay the people you ask for money.
- We can set up a cost recovery center! (or faculty are willing to part with money...)
 - This just doesn't work, at least not entirely. Academic centers that tried this are gone. Ask SUN how it's going outside universities...
- I won an MRI, and the hardware is the hard part!
 - Hardware money is relatively easy, and is actually not the biggest cost. Facilities cost for 20KW racks are huge, and personnel costs forever.



IRA A. FULTON SCHOOL OF ENGINEERING
HIGH PERFORMANCE COMPUTING

Some (poorly researched) Examples

- Research Office Funding: TACC, Utah
 - In one case, the ROI is huge. In the other, direct support to a tremendously successful vis group, and a long tradition of great service to faculty
- Central IT Funding: Oklahoma, Maryland, U. of Arizona
 - In some places, this has been successful. In other, impact is limited to a few faculty, or resources are inadequate to do meaningful work.
- Private Funding: BYU
 - One primary donor; Lots of machines, personnel, facilities are a huge challenge



IRA A. FULTON SCHOOL OF ENGINEERING
HIGH PERFORMANCE COMPUTING

Some (poorly researched) Examples

- State Investment:
 - OSC (Ohio), LSU (CCT, LONI), Indiana
 - In Ohio, focus on many state institutions, connections to federal facilities
 - In Louisiana, focus on economic development
- In all 3 cases, state investment has totalled >\$50M
 - Requires careful advocacy
 - Not likely to be replicated everywhere
- Through investment in resources, all 3 have acquired Teragrid funding (but this will go away).
- Substantial leverage of other grant funding



IRA A. FULTON SCHOOL OF ENGINEERING
HIGH PERFORMANCE COMPUTING

The Good News: Leverage

- HPC / Research Computing can affect all these areas...
 - Research office is already investing in HPC through start-up packages (on individual faculty clusters... in our case, >\$400k/yr)
 - IT is incurring support costs, one way or another
 - Someone is covering facilities costs (find out who)
 - HPC can drive research across a broad range of areas
 - HPC competitiveness focus lately makes ED an easier argument
- It's *not* just another piece of research infrastructure.



IRA A. FULTON SCHOOL OF ENGINEERING
HIGH PERFORMANCE COMPUTING

The ASU Experience

- Begun with endowment funds
- Housed in the Fulton School of Engineering
- Construction planning began 9/04
- Operations began 9/05



IRA A. FULTON SCHOOL OF ENGINEERING
HIGH PERFORMANCE COMPUTING

Part of ASU CyberInfrastructure Portfolio

- Decision Theater - Flagship Visualization facility
(and close partner with HPCI)
- Fulton HPC Initiative
- UTO -> Campuswide storage system
- National Lambda Rail / CENIC connectivity



IRA A. FULTON SCHOOL OF ENGINEERING
HIGH PERFORMANCE COMPUTING

HPCI Services

- Hard:
 - Cycles
 - Storage {long term archive, high speed scratch}
 - Network (sort of)
- Human:
 - Application / Consulting / Training Services
 - Proposal support {Federal Grant; Supercomputer Center Allocations}
 - Teragrid
 - Visualization
- Collaborative Research



IRA A. FULTON SCHOOL OF ENGINEERING
HIGH PERFORMANCE COMPUTING

Business Model

- Substantial costs to run this center... computers are a fairly small part
 - Maintenance, license agreements, power, disk storage, backups, etc.
 - Staff - sys admin, security, updates, support, programming, debugging...
- Can't use endowment funds forever...
- Costs must be recovered somehow
- Investment to date:
 - Physical Infrastructure ~\$1M, Equipment ~\$2M, People ~\$0.6M.
 - FSE Investment ~\$2M



IRA A. FULTON SCHOOL OF ENGINEERING
HIGH PERFORMANCE COMPUTING

Business Model

- “Hard” Services are a co-investment

Core funds provide people, infrastructure; faculty / projects provide hardware capacity (examples to follow...)

- Human services require more cost recovery

Applications / Consulting must pay for itself

Grant services pay for themselves in more obtuse ways

Training is a core activity



IRA A. FULTON SCHOOL OF ENGINEERING
HIGH PERFORMANCE COMPUTING

Allocations - Cycles

- Three - tiered allocation model:
 - A minimum level of service for free
 - Additional service via proposal to faculty-run allocation committee
 - Services beyond that on partial cost recovery basis
- Storage is handled in similar fashion

Business Model - Core Funds vs. Project Funds

- I have no doubt I could run this center as 100% cost recovery... the first thing I would do is kick most of the faculty off my systems.
- Core funds are required for service mission
- Annual Operation Costs:
 - \$1M Core
 - \$1M HPC Research funds
 - \$500k cost recovery (service revenue)



IRA A. FULTON SCHOOL OF ENGINEERING
HIGH PERFORMANCE COMPUTING

Funding Model

- Core funds: \$1M
- University Investment:
 - UTO (IT), Research Office, contribute \$200k each annually (40% of total)
- Three Academic units investing \$200k each:
 - Fulton School of Engineering
 - Liberal Arts and Science
 - Life Sciences



IRA A. FULTON SCHOOL OF ENGINEERING
HIGH PERFORMANCE COMPUTING

Remaining Unsolved Issues

- Service vs. Equipment and **Indirect Cost**
- Portfolio Balance
- Protecting Core Funding

Service vs. Capital

- Technically, most of what I just described is illegal to sell without charging indirect costs
- Typically, the funds I'm targeting are equipment funds, which are exempt from indirect cost.
- This represents a significant but surmountable barrier.

Portfolio Balance

- I've won \$1.8M in grants this past summer
- This behavior is encouraged by the administration
- Each new project requires time
- My systems scale, my days do not
- Currently about 40 open projects, including many internal efforts.
- Staff is growing, but each new project requires project-focused staff
- *Protecting core staff and service focus is an ongoing challenge*



IRA A. FULTON SCHOOL OF ENGINEERING
HIGH PERFORMANCE COMPUTING

Take Aways, and What's Next

- Centers are good, but are partial solutions
- Campus funding is politically complex, but doable.
- Centralizing HPC on the campus is worth doing.
- We need more discussion:
 - I'd like to hold a workshop on this...
 - Start a more constructive dialog



IRA A. FULTON SCHOOL OF ENGINEERING
HIGH PERFORMANCE COMPUTING

Discussion