Global Grid Forum

Promoting and Standardizing Grid Computing

When Worlds Collide

Andrew Grimshaw
University of Virginia
October, 2005

Name that tune

"When worlds collide, said George Pal to his bride,
we're going to give you some terrible thrills."

The story of Dr. Frankenfurter, Brad and Janet.

Does this remind you of any of your experiences?

There are many worlds that come together in Grid

- Real world
- IT management world
- Computer/computational science world
- Worlds of various scientific and engineering disciplines
- Vendor world
- Commercial worlds

Each of these worlds perceives Grid differently – and wants different things from Grid.

Real world

- Real world views Grid, if at all, through press releases and low content press pieces.
- Now would be my opportunity to take some cheap shots at some of the vendors …. But I won't.
- The real world does not want to know about Grid any more than they want to know about how the pluming works – and what happens when they flush

IT Management

- Is risk averse
- Have enough problems on their hands
- People are their most expensive resource
- Are the ones blamed if something goes wrong

Computer/computational Science

- Always looking for new problems
- Want to do cool things – tend to be very technology focused
- Are looking for the abstract form of the problem
- Are rewarded for publications
- First 90% is the most fun
- Don’t want to be IT staff
- Want to help

Science and Engineering Disciplines

- Want to solve their problem – computers are a means not an end
- Would rather not know about their computers – it is surprising how many don't know "the basics"
- May think primarily of short term – "I need more memory"
- May not think beyond the immediate
- Energy barriers
- May not know where they are going or what they are trying to find
- Often work on significantly different timelines

Vendor World

- View Grids as a means to an end
- Objective is to make money – don't forget that!
- Many/most companies focus on quarters – longer term projects are difficult for all but the larger companies
- Are interested in activities that improve their competitive advantage
- Want to make money!

Commercial World

- Want solutions to problems – not new cool technology
- Want to decrease costs and improve productivity
- Want to know the ROI case – and must be convinced that it is true – energy barrier
- Grid is inherently a corporate wide sale –
- CFO's have a lot of power

Many different world-views

- Complex interplay of interests
- Simplicity is key – nobody wants a major change in their
  - Code
  - IT infrastructure/
  - operational methodology
  - cost structure
  - …

Life Sciences

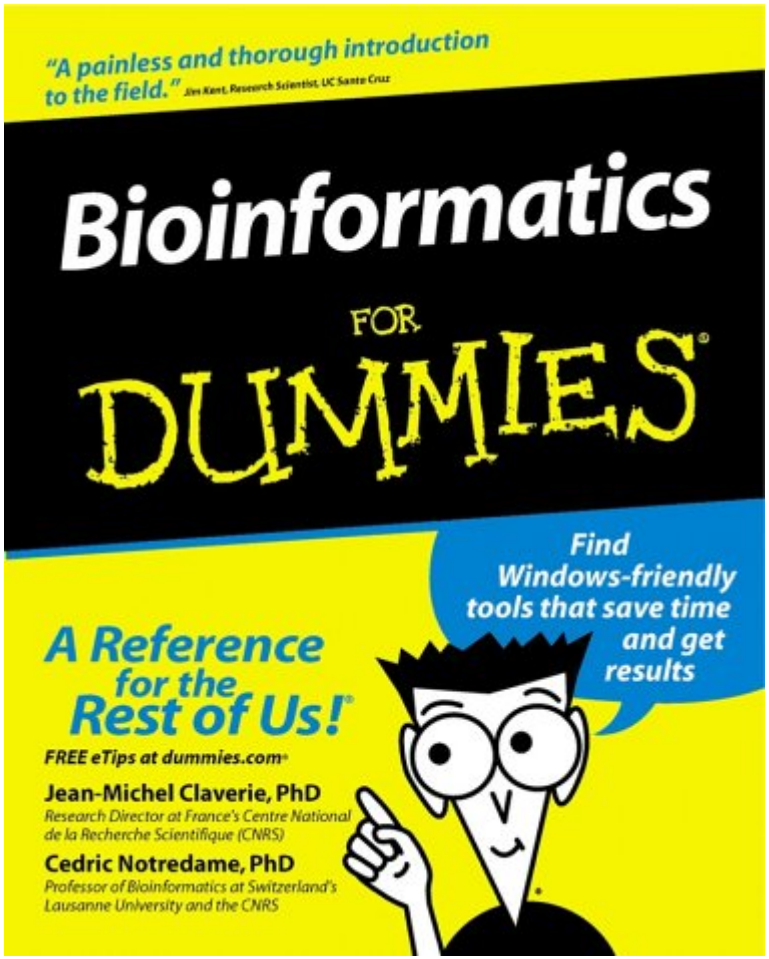
In ten years the world will be very different.

Think back ten years.

- No web
- Wide-spread internet was new
- Human Genome Project still far from completion
- Science (biology) done primarily in individual labs
- Medicine in the US a fragmented industry

Today

- Billions a year in e-commerce
- Internet everywhere
  - Broadband to your home
  - Wireless becoming pervasive
- Pervasive device are proliferating – motes
- Science still done mostly in individual labs
- Sequencing of organisms a daily event. Bioinformatics has hit the main stream
- Medicine is still a fragmented industry


"A painless and thorough introduction to the field." Jim Kent, Research Scientist, UC Santa Cruz
Bioinformatics
FOR
DUMMIES
Find Windows-friendly tools that save time and get results
A Reference for the Rest of Us!
FREE eTips at dummies.com
Jean-Michel Claverie, PhD
Research Director at France's Centre National de la Recherche Scientifique (CNRS)
Cedric Notredame, PhD
Professor of Bioinformatics at Switzerland's Lausanne University and the CNRS

Tomorrow

- $1000/sequnce for humans – becomes standard clinical practice
- Global interconnected networks – grids
  - Provide transparent, secure, access to data, applications, and on-demand compute.
- Research using not just your data, but all *trusted* data, not just your applications, but any *trusted* application.

Tomorrow - Clinical

- Unified clinical records
- Data mining of integrated clinical and scientific databases

What this will facilitate

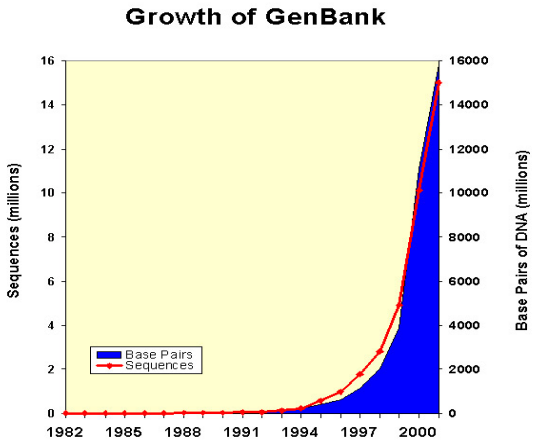
- Discovery of new diagnostics and predictors for prescriptive medicine
- Discovery of new treatments to complex genetic diseases
- Push back expected lifetimes and improve quality of life
- Deliver better health care by identifying problems earlier
- Wide spread epidemiological research

There are a number of “catches”

- So much data!
- So many organizations with so little trust!
- So much complexity!

Increasing Cost and Complexity of Life Science Data-sharing

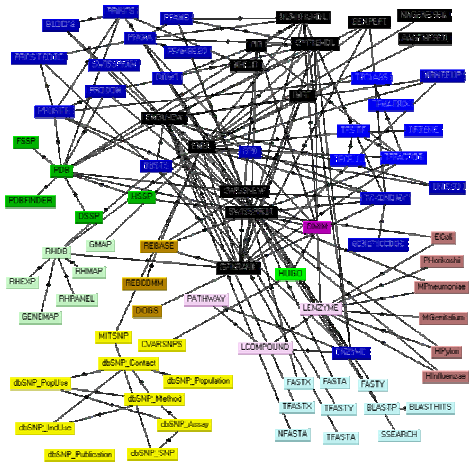
- Over 330 public Life Sciences DB's
- 8x growth in genomics data, last 18 months[1] … just the beginning …
  - Proteomics data: 1,000x multiplier[2]
  - Glycomics, new small molecule efforts
- Increasing scope of data diversity
  - Annotations (interactions)
  - Organism-specific (mouse, human)
  - Molecule-specific (protein, sugar)
  - Data-type-specific (gene expression)
- Complex data interrelationships
- Redundant effort to download/maintain data
- Data coherence challenges
- Difficulties in sharing data across (and among) organizations



Genbank growth has continued to trend sharply upwards, as have many other classes of biological data

Increasingly-complex interrelationships between biological research databases today

(LION graphic)



[1] **Source: TimeLogic estimates.** [2] **Frank Gleeson, CEO, MDS Proteomics**

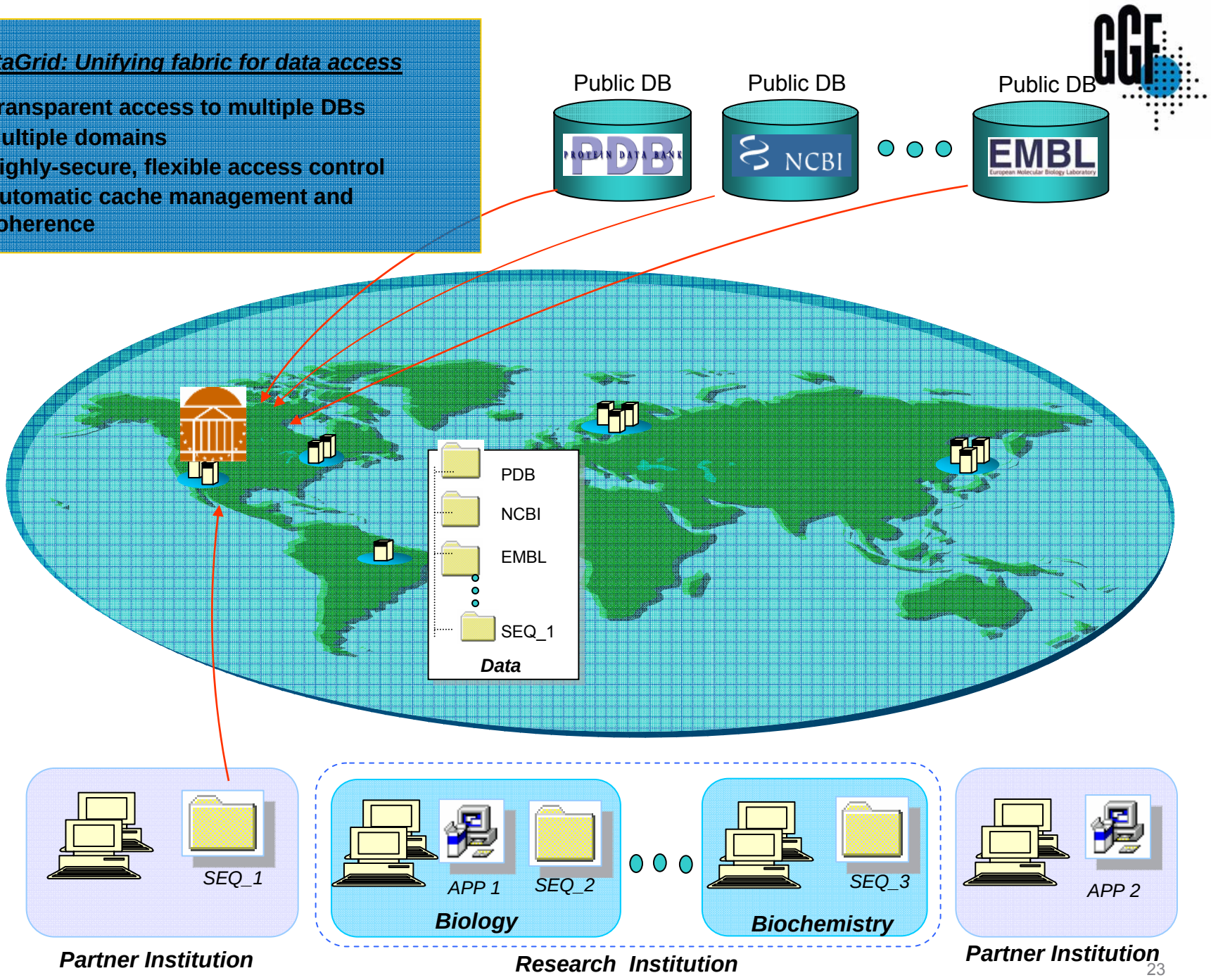
This is where grids enter the picture

Global Bio Grid at Virginia

- Bring Grid capabilities to both life sciences research and to clinical integration of patient records.

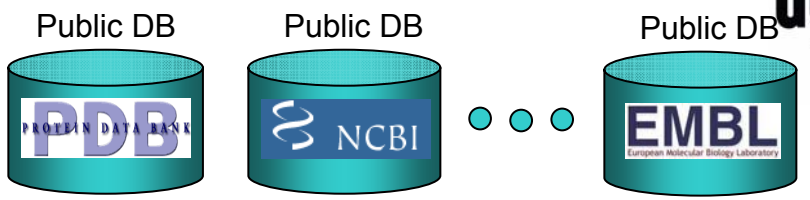
DataGrid: Unifying fabric for data access

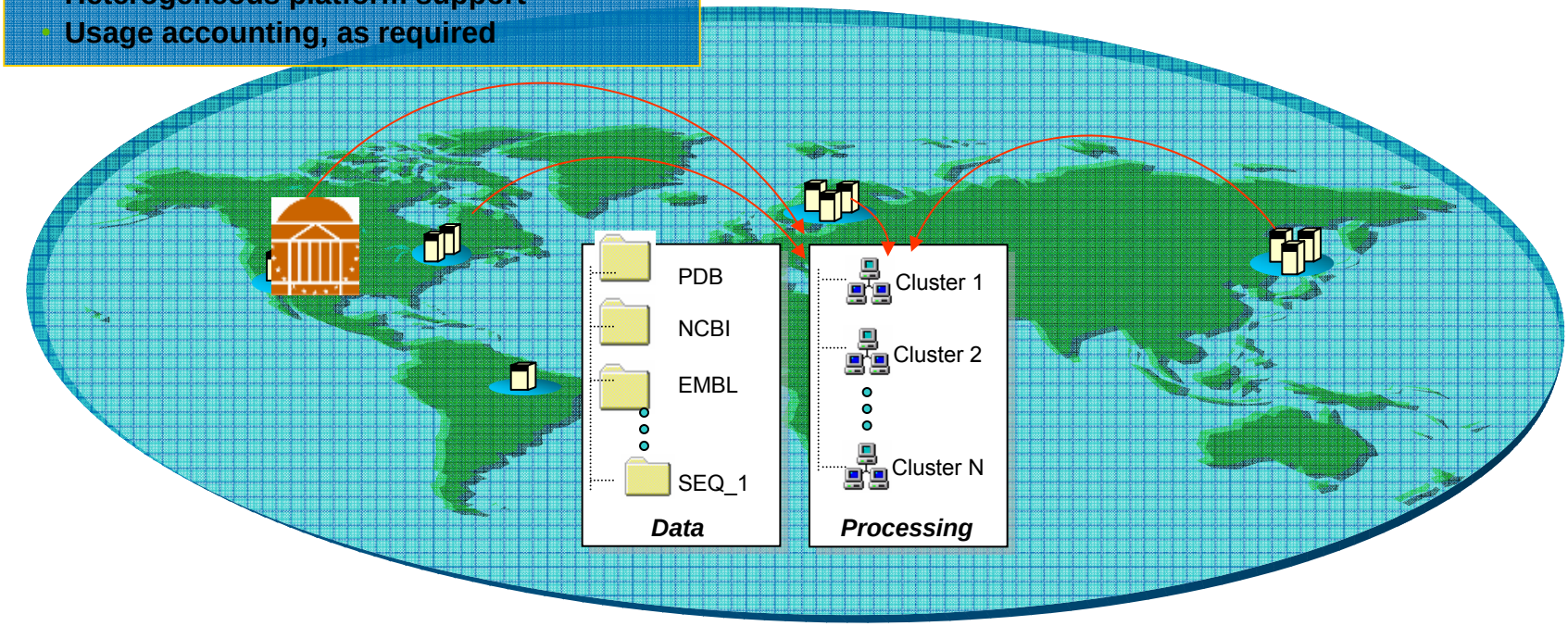
- **Transparent access to multiple DBs**
- **Multiple domains**
- **Highly-secure, flexible access control**
- **Automatic cache management and coherence**

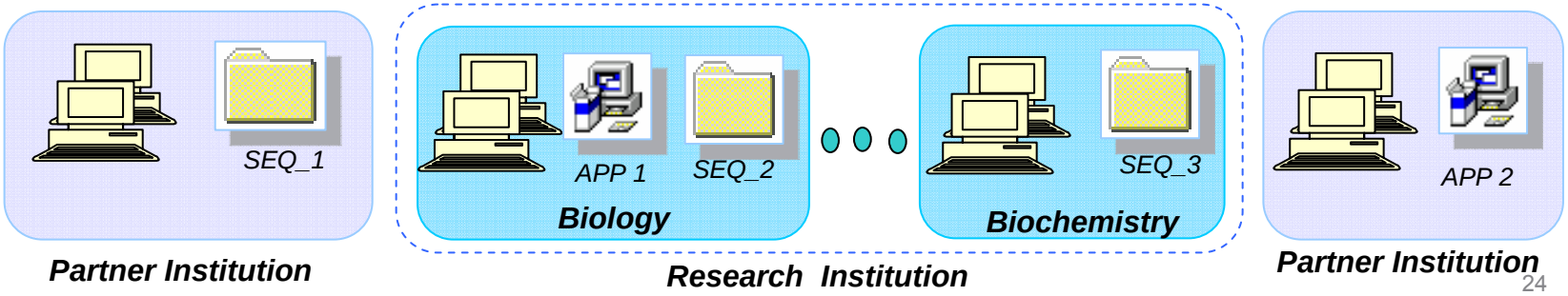


Compute Grid: Shared access to processing

- **Flexible, location-independent access to virtually unlimited processing, on-demand**
- **Scheduling, usage, management policies**
- **System detects, recovers from job failures**
- **Heterogeneous platform support**
- **Usage accounting, as required**

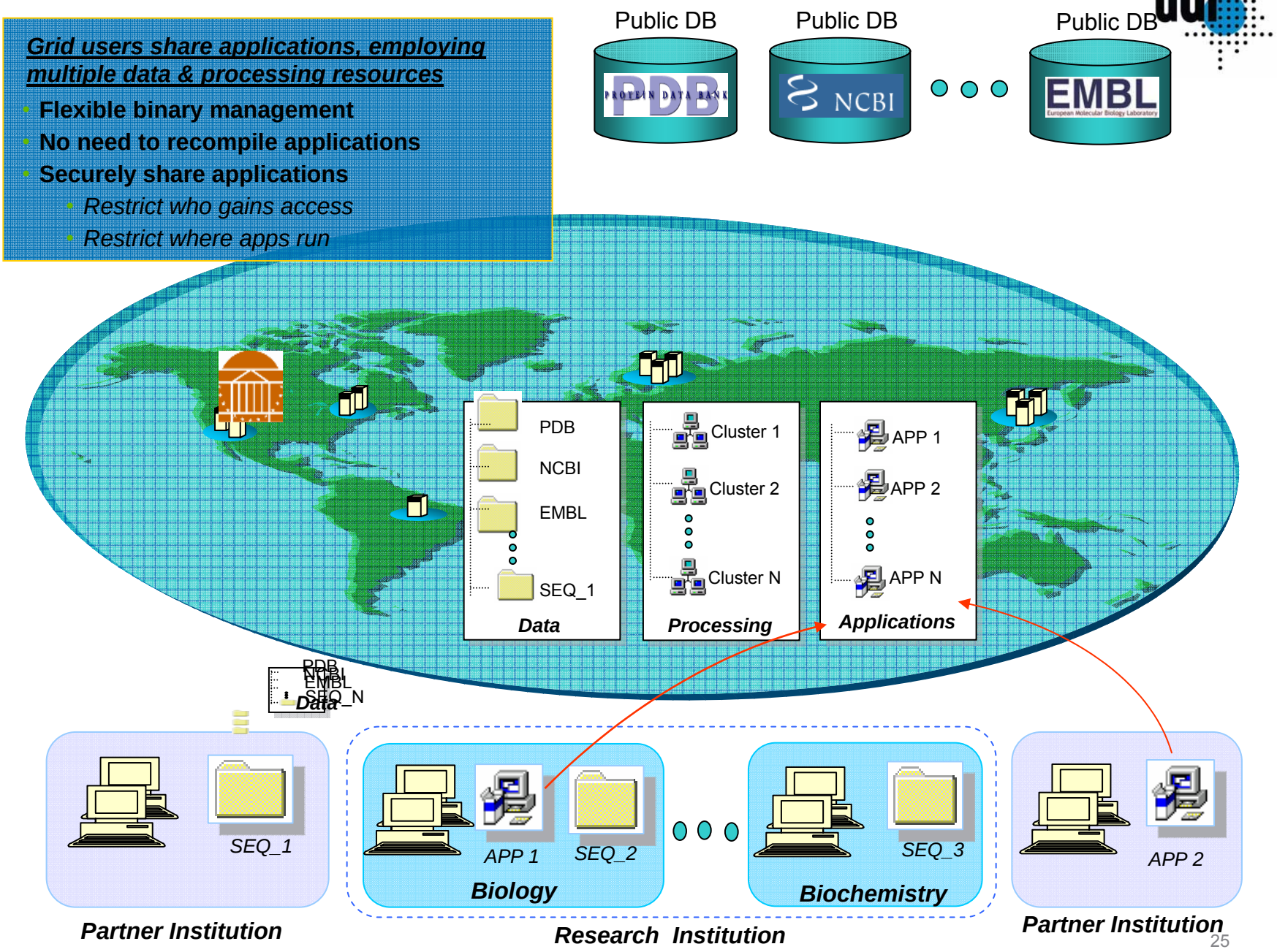






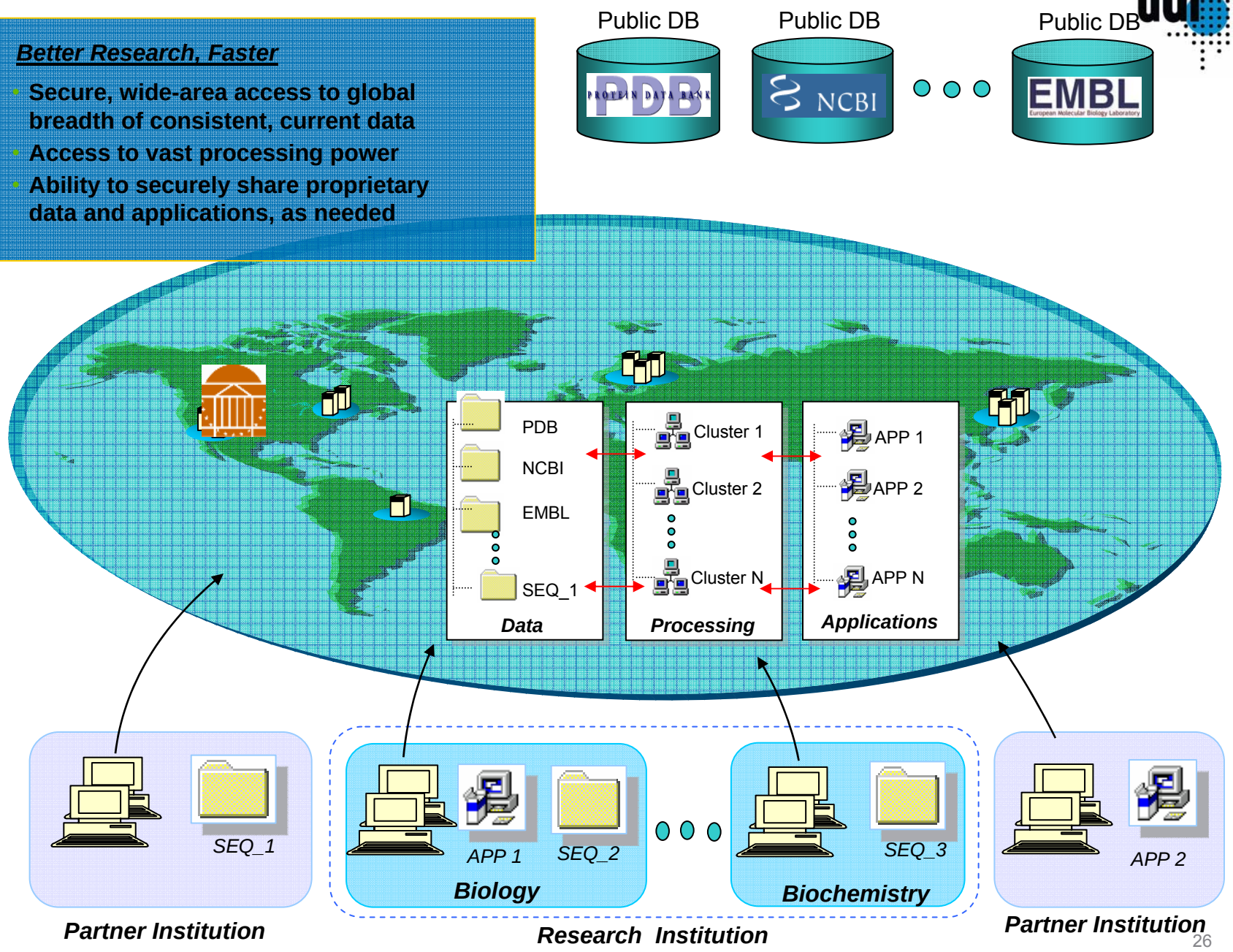
Grid users share applications, employing multiple data & processing resources

- **Flexible binary management**
- **No need to recompile applications**
- **Securely share applications**
  - *Restrict who gains access*
  - *Restrict where apps run*



Better Research, Faster

- **Secure, wide-area access to global breadth of consistent, current data**
- **Access to vast processing power**
- **Ability to securely share proprietary data and applications, as needed**



Case study – Large Central Virginia Teaching Hospital

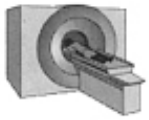
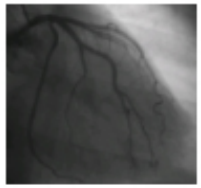
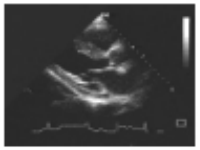
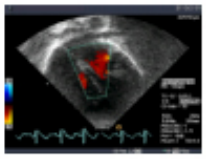
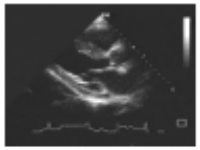
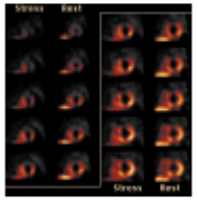
- Medicine is far behind the rest of the world in the use of IT in their “business”
  - Information in silo’s
  - Don’t show up in the emergency room expecting them to know about your by-pass you had last week – even at the same hospital!
- Want to integrate “extended hospital”
- Research opportunities abound
  - Sequencing, protein expression becoming less and less expensive
  - Bio marker research. Combine clinical data (including outcomes) and genomic/proteiomic data
  - Imaging

The Players

- Think western
  - “The Good, The Bad, and The Ugly”
- Patients – who want to live, have a better quality of life, have their confidential records kept confidential, though they also want better cures in the future.
- MD’s who want to provide better care for their patients both in the short term via Clinical Data Integration (CDI) and real-time MRI; and in the long term via discovery of better clinical rules and integration of clinical records and genomic and proteiomic data.
- Bioengineers/researchers who want to develop new techniques for
  - modeling of tissue growth
  - 3D real-time MRI
  - Biomarkers
  - ……
- Myriad IT managers in the hospital – each of which has their own silo’s of information and is responsible to different department heads

Image Management and Interface – Scope

Combining Imaging Data and Metadata from Multiple Sources with Structured Reporting in Real Time and at the Point of Care

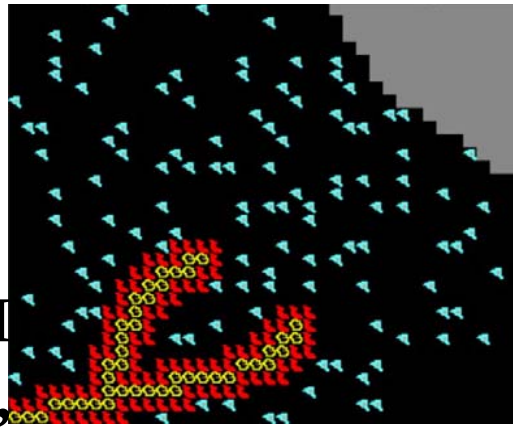
Cardiovascular PACS							
Clinical							Research
							
Cardiac Cath and Electrophysiology	*Adult Echo*	*Peds Echo*	*Vascular Lab*	*Nuclear / Ex. Stress*	*Radiology PACS*	*Systems Integration*	*Emerging Technologies (cardiac MR, CT angiography)*
							
Digital video	*Digital video*	*Digital video*	*Digital video*	*Single frame images*	*Single frame images*	*HL7 interface*	*Digital video*
Vendor Integration							
Philips Integris Philips Allura Siemens Coroskop GE Innova	Philips Sonos Philips HDI Siemens Sequoia Toshiba PowerVision	Siemens Sequoia Siemens Cypress	Philips HDI	GE Infinia Philips Prism	DICOM Query and Retrieve Interface to Radiology PACS (Kodak)	SMS Patient Reg. Witt Physiomonitoring PedCath 7 KinetDx (peds echo) Datacheck (vascular) Tracemaster CAOS Cardiac IS IDX CareCast	Siemens Coroskop (XA) Siemens Sonata (MR)

Predictive Systems Biology

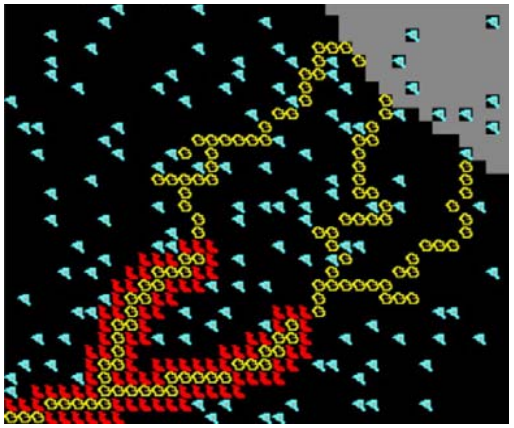
Computer Simulation of Arteriolar Pattern Formation

Biomedical Engineering, Cell Biology (SOM) Computer Science, Biology (A&S)

A. Arteriole: Endothelial cells covered by smooth muscle cells.

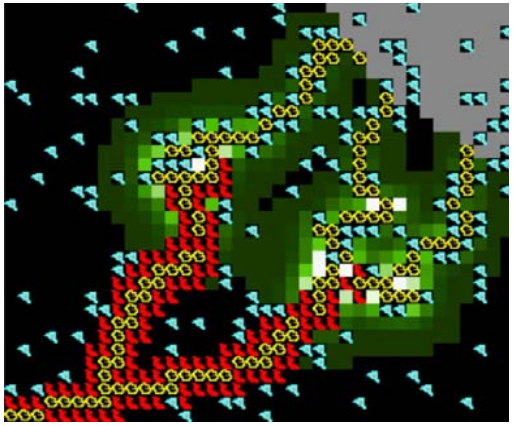


B. Endothelial cells sprout toward a VEGF source.



C. Capillaries release PDGF-B.

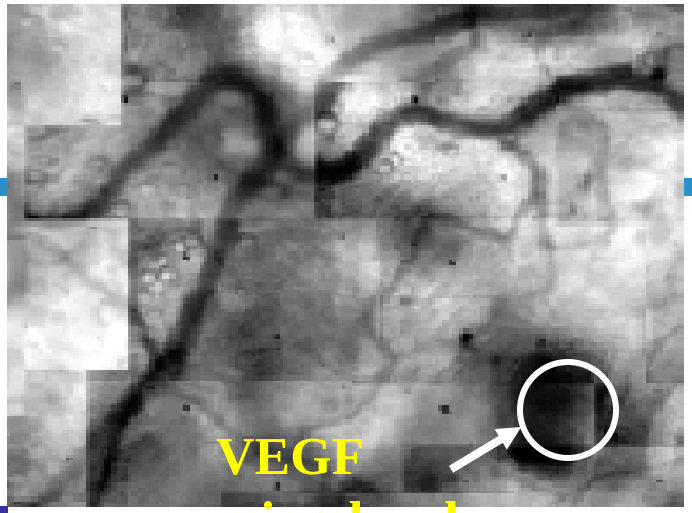
D. Fibroblasts migrate to capillaries and differentiate into smooth muscle cells.

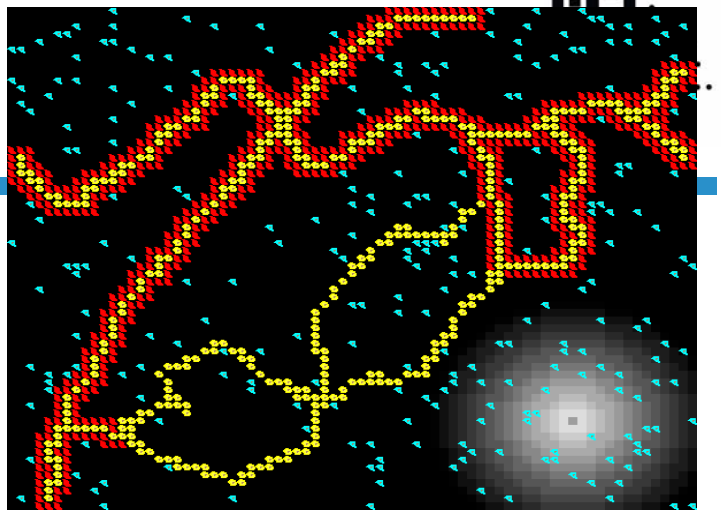


Smooth Muscle Cell Fibroblast Endothelial Cell

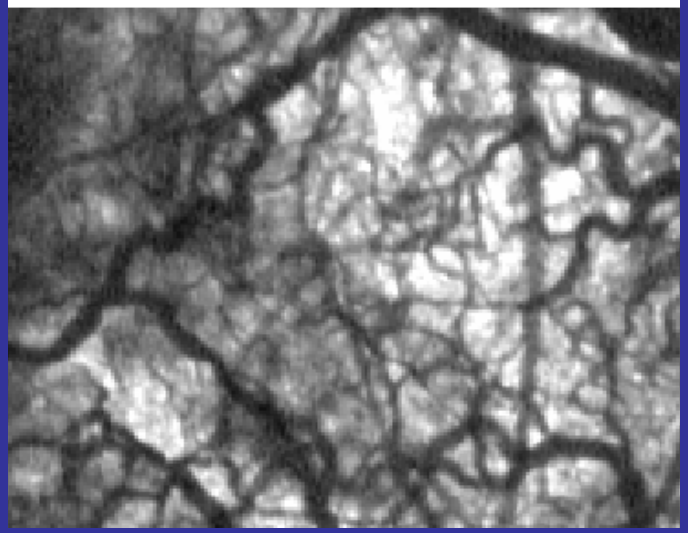
VEGF Source PDGF-B Gradient

Day 1





Day 14

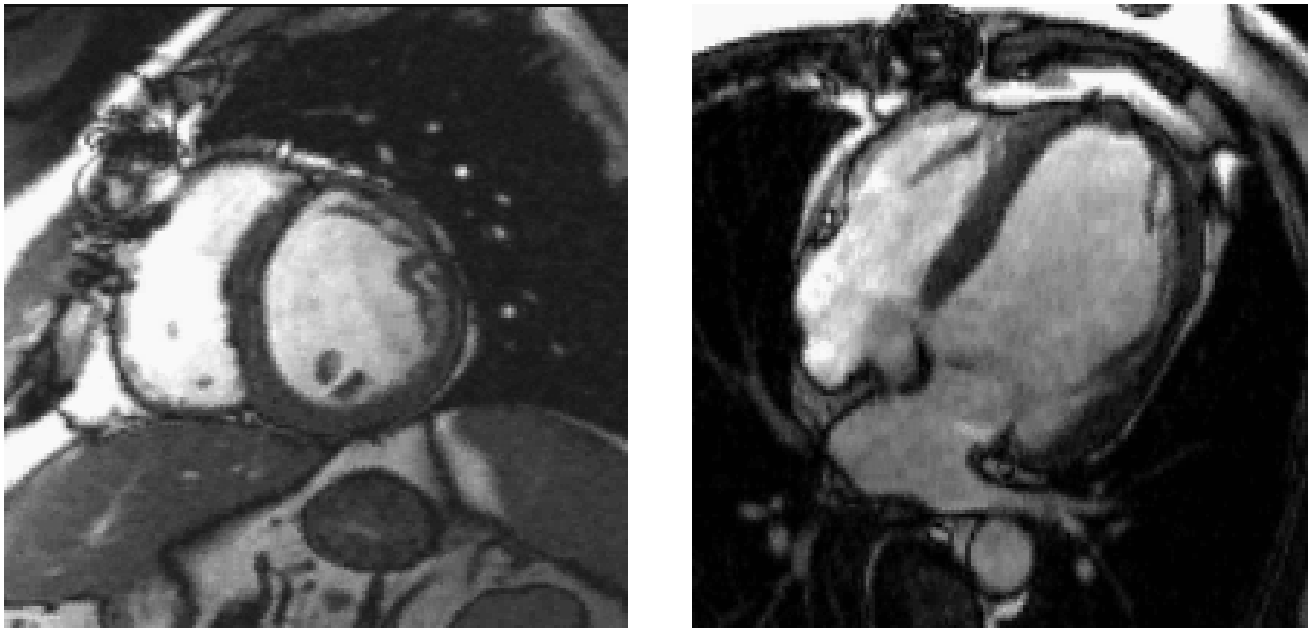


IN VIVO

IN SILICO

Identifying new targets through computational mo

LV structure and function



Steady state free precession imaging

Why a Real-Time Stress Test?

- No dependence on breath-holding
- No dependence on cardiac gating
- Arrhythmias not a problem
- Rapid coverage of the heart
- Familiar to cardiologists from echocardiography
- But – technically challenging to maintain spatial resolution with improved temporal resolution

The players - continued

- The Dean of the School of Medicine who wants the institution to climb in the rankings and wants there to be more money in the system so he can get more overhead. The Dean (and the rest of the hospital) are used to collaborative research deals with pharma's where tens of millions of dollars change hands.
- The head of the hospital who wants to keep costs down no matter what and not get sent to jail for a HIPAA violation or have the institution go bankrupt in court. His incentive structure is money – not patient outcomes. Must see proven value.
- The hospital CIO who started a data warehouse project seven years ago ….. And has bet the farm on success

The players - continued

- Payers – insurance companies and the government want costs contained – but also will pay more for provably better outcomes.
- Big vendor. Their objective is to make money. Wants to move into a new, huge market (16% of US GDP in health care, significant portion of that on records management). Wants to leverage its technologies developed in other commercial spaces. Wants to sell gear and services to hospital. “Pay us and we’ll make the vision real at your hospital.” Not used to paying for large collaborative projects.
- Computer scientists. Naturally the good guys with nothing but the purist of motives. They have hammers – are looking for nails. Want publications, money to pay graduate students and staff, and access to the best toys. Also have a complex .. “we’re not just programmers”.

The players - continued

- The US federal government
  - HIPAA provides for significant civil liability and criminal liability on un-authorized disclosure of confidential information.
  - “De-identification” of data is legal standard
    - BUT – it is recognized to not be technically sufficient
- Hospital Internal Review Board
  - Must authorize any project/action that involves human subjects or their data
- Hospital/University lawyers
  - Estimate that legal liability could run to US $3M per non-authorized disclosure.
  - Hospital has records for 100,000+ patients
  - You do the math!
- Big vendor lawyers

The Plot

- Everyone proclaims shared common vision of personalized medicine, predictive biology, and fully integrated health care.
  - Life is good.
- Easy stuff (things not requiring lawyers and lots of money) are on-line
  - Campus-wide compute grid using both cluster and desktop resources
  - Data grid linking data sources
  - Biomarkers applications running
- Still have a long way to go

Summary

- There are many different players in a successful Grid solution – all must be co-opted.
- Grid is as much sociology as technology – it is a paradigm shift that forces people to work together to get the benefits
- Changing the *status quo* is difficult
- Being a Grid computer scientist is great fun because I get to interact and learn from so many different people in so many different domains.