Many Different Types of GRIDS ?

GRID Technology is not just about CPU or Data

its about ALL SERVICES

Cycle Scavenging GRIDS – deskt[illegible]

Super[illegible]ces

[illegible] (e.g. distrib[illegible]ers, simulations, [illegible]eractive respon[illegible]

Data GRI[illegible] large amounts of distributed data, data tends to [illegible]omputed)

DataGRI[illegible] or DatabaseGRID, (Genomics, Ensembl, Proteins etc.)

Service GRID (Astronomical combined data, BL[illegible] on genome, protein databases)

Collaborativ[illegible] GRI[illegible] V[illegible]…steering …

Integratio[illegible]ed heterogeneous dbs)

…other GR[illegible] classifications…

An Evolving and Wider Vision

Thinking about The GRID is evolving to encompass not just computational and data resources but to…

The decision maker, researcher, Doctor, surgeon…

In his own Virtual Organisation of collaborators

is provided with his own virtual col-laboratory, virtual environment, workbench environment…

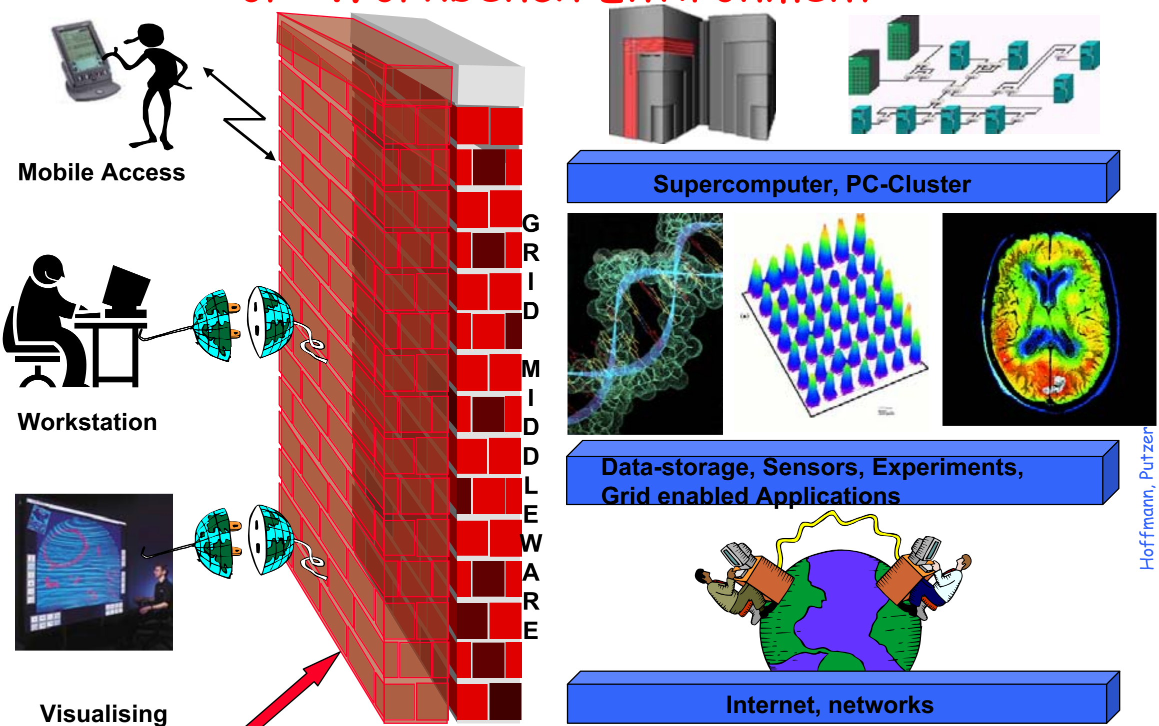
In which computers and technology make available to him in a transparent and easy fashion, independent of location and time…

All services that he/she needs for the best decision making…

which may be information of all types from databases (including e.g. medical records…), compute power for applications (e.g. modelling), access to colleagues for consultation,

And the GRID is just an enabling technology in this vision

The User connects to his "Virtual Laboratory" or "Workbench Environment"



Integrating framework middleware

Examples of Different GRIDS

Simple Cycle-stealing GRIDS

SETI@home

Oxford University/United Devices: screening Anthrax toxin, Smallpox, certain cancers…

Pharmaceutical companies:

- Novartis 2700 computers used to discover one most promising drug candidate
- Plans to extend to much wider network

Ask a question you thought you could never ask

Reduce a "batch" problem to an interactive question!

A major change in productive working!

"Crunch GRIDS"

Aim: to be able to ask questions that did not make sense before because they took too much time

Reduce a "batch" problem to an interactive question!

Computer Farms
and "Cycle Harvesting"

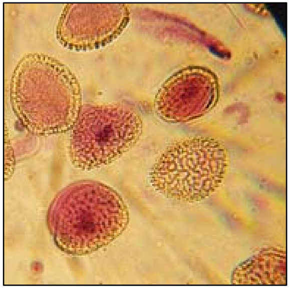
The example of NOVARTIS

Pharma Grids – Key to Pharmaceutical Innovation?

René Ziegler, PhD
Head of Global IT Management
Novartis Pharma AG
CH-4002 Basel, Switzerland
rene.ziegler@pharma.novartis.com

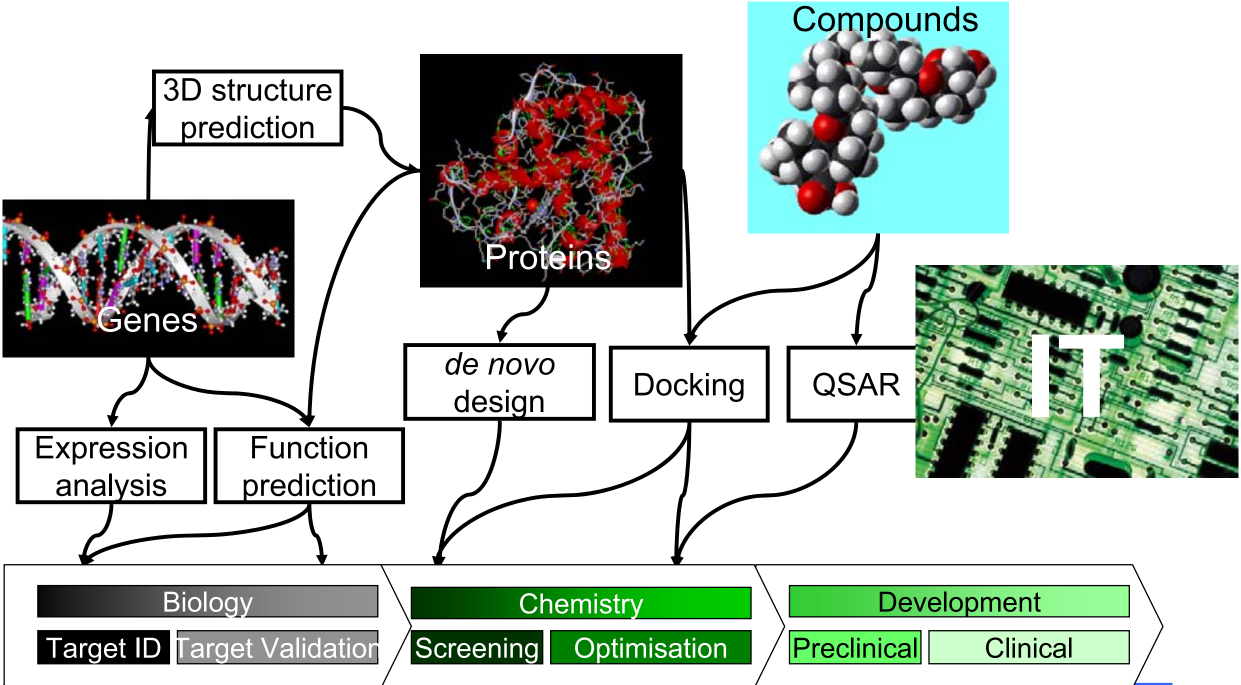
The Evolving *in silico* Science Platform

Evolution of the Pharmaceutical Industry:

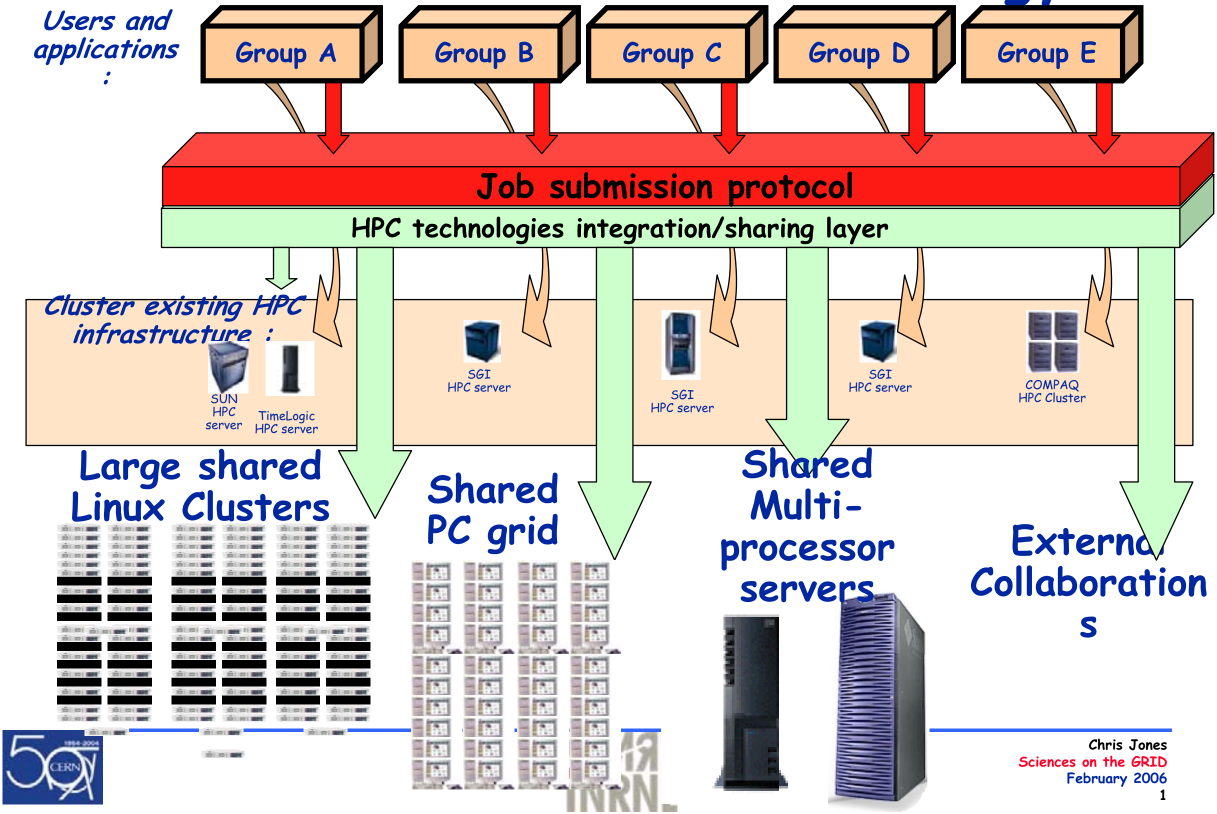
In Vivo	*In Vitro*	*In Silico*
		

[Time]

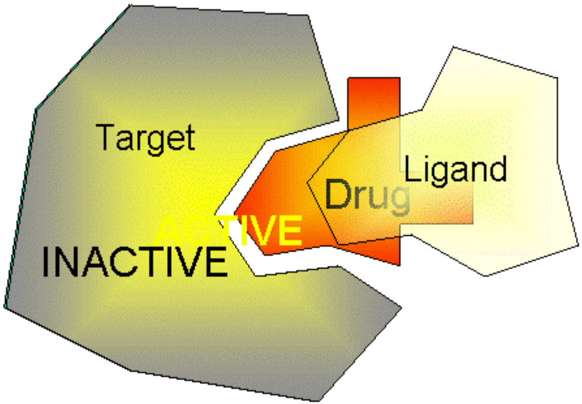
The Contribution of IT to Drug Discovery is Increasing



Novartis Has Implemented a Globally Coherent Pharma Grid Strategy

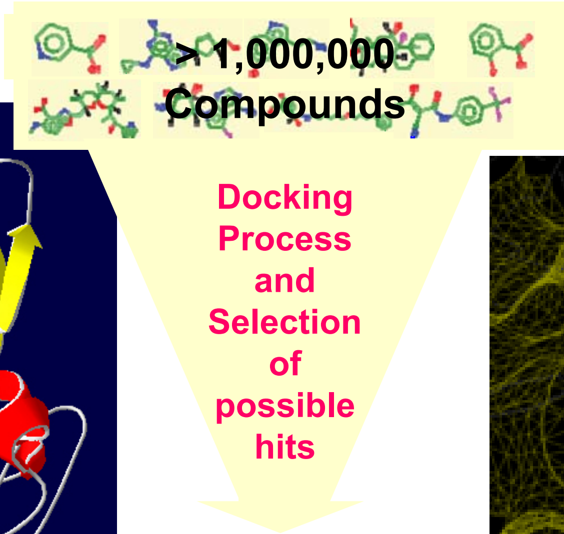


Influencing Bio-molecular Processes

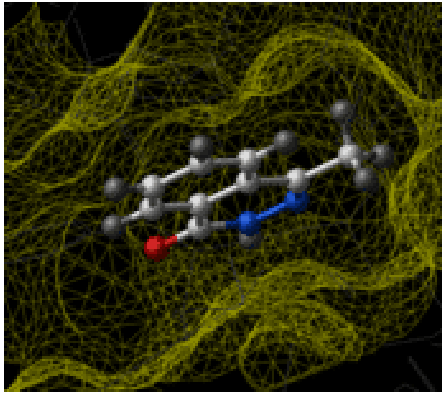


Target = enzyme, receptor, nucleic acid, ...
Ligand = substrate, hormone, other messenger, ...

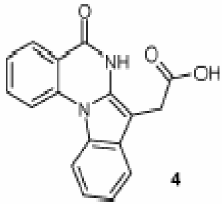
Virtual Screening by *in silico* Docking



< 100
Compounds



Results



4

Conclusion

We have identified a 7-substitued indoloquinazoline compound as a novel inhibitor of protein kinase CK2 by **virtual screening of 400 000 compounds**, of which a dozen were selected for actual testing in a biochemical assay. The compound inhibits the enzymatic activity of CK2 with an IC50 value of 80 nM, making it **the most potent inhibitor of this enzyme ever reported**. Its high potency, associated with **high selectivity**, provides a valuable tool for the study of the biological function of CK2.

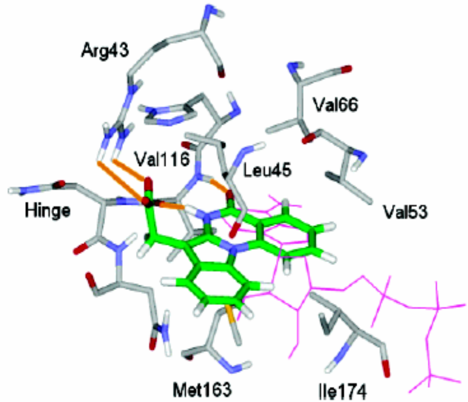


Figure 3. Relative binding modes of compound **4** (green) and the ATP analogue extracted from the crystallographic template structure 1DAW (magenta) in the human CK2α active site. The hydrogen bonds formed with Arg 43 and Val 116 are indicated as orange lines.

Table 2. Selectivity Profile of Compound **4** in Terms of % of Inhibition of the Catalytic Activity of the Kinases at a Compound Concentration of 10 μM

CK2	c-Abl	HER-1	HER-2	KDR	Flt-3	IGF-1R	c-Raf-1	PDGFR-ß	c-Kit	Flt-4	Flt-1	Tek	PKA	c-src	CDK1	PKB	PDK1	Ins-R	FGFR-1	c-Met
97	25	12	29	54	41	44	30	11	31	30	21	36	51	0	51	30	9	23	32	40

Virtual Docking Accelerates the Docking Process at Negligible Additional Cost

Task: DOCK, ~320,000 molecules

Virtual docking of compounds from the Novartis Library into the 3D structure of a protein (target)

Elapsed time Hrs.	55,794	547	6
Elapsed time days	2,325	23	
Elapsed time years	6.4		
Devices in Grid:		561	1200


René Ziegler, PhD
Head of Global IT Management
Novartis Pharma AG
CH-4002 Basel, Switzerland
rene.ziegler@pharma.novartis.com

"Crunch GRIDS"

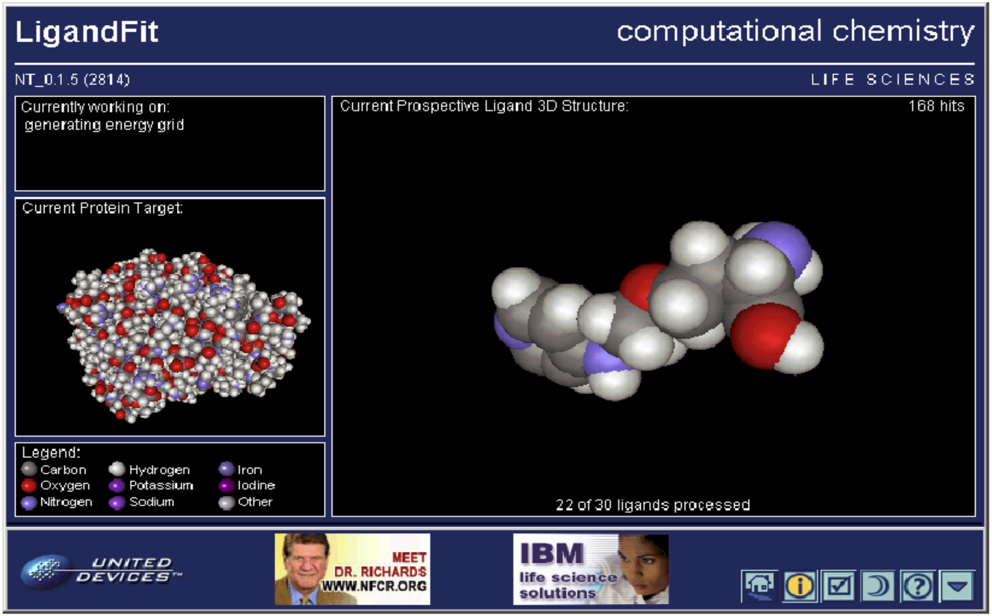
"Cycle Scavenging" with screen saver technology
(as used by SETI@Home)

Example of Oxford University, United Devices, IBM, Accelrys
www.GRID.org

In silico screening to select best molecules for *in vitro* bio-assay.

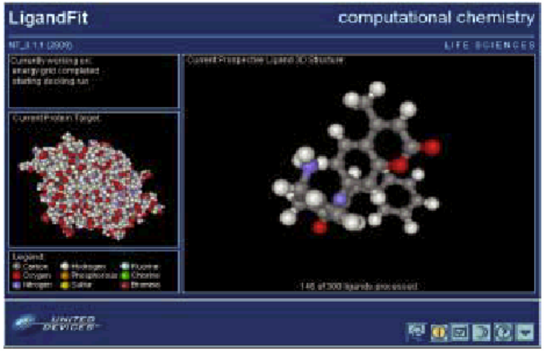
- Cancer
- Anthrax Toxin
- Smallpox

Grid.org


LigandFit
computational chemistry
NT_0.1.5 (2814)
LIFE SCIENCES
Currently working on:
generating energy grid
Current Protein Target:
Current Prospective Ligand 3D Structure:
168 hits
22 of 30 ligands processed
Legend:
Carbon
Hydrogen
Iron
Oxygen
Potassium
Iodine
Nitrogen
Sodium
Other
UNITED DEVICES
MEET DR. RICHARDS
WWW.NFCR.ORG
IBM
life science solutions

The Cancer Project

- **Goal**
  - Exhaustive screen of 12 targets identified in multiple cancers using the world's largest molecular library
- **Challenge**
  - Massive computational power was required to realize Oxford's vision
- **Solution Highlights**
  - Built a working Public Grid dedicated to Life Science research
  - Accomplished the initial stage of the project in less than 12 months
- **Results**
  - Over 2.5M nodes from around the world are part of Grid.org
  - Total CPU time for project was over 190,000 years
  - Results published in Nature Drug Discovery



But these are simple GRIDS

The Vision in Life Sciences, and in many other sciences, is to use the GRID NOT just for cycle harvesting or scavenging, but to extend human capabilities with e.g.

Service GRIDS

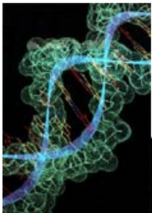
Information/Knowledge GRIDS

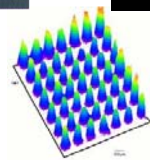
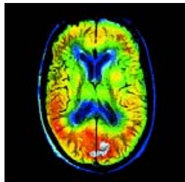
More informed decisions

Whilst many services and much information are potentially available to the decision maker today,

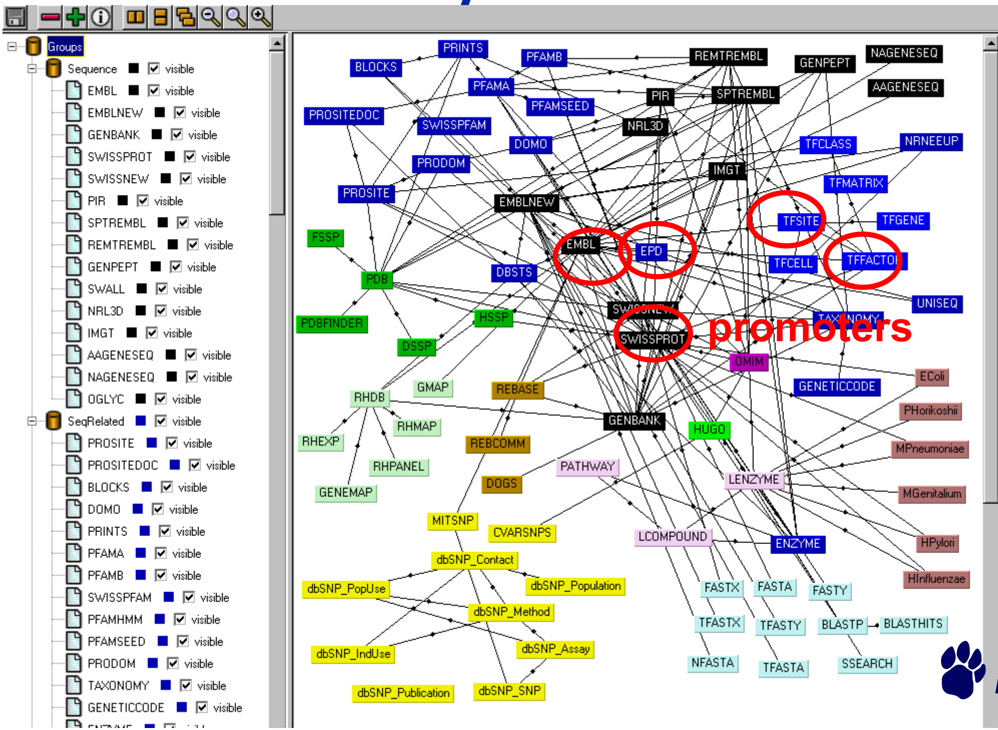
in practice there are so many barriers that such access and integration are not even considered

The vision is hence of the research worker or doctor able to make much more informed decisions because so many new services and information are available to support the decision.



'Real world objects' are distributed over many databanks



LION

Virtual Observatories

An example of a SERVICE GRID

Astrophysicists have images of objects in many different wavelengths, in different distributed heterogeneous databases:

Optical

Infrared

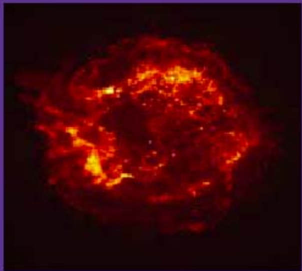
X-ray

Radio

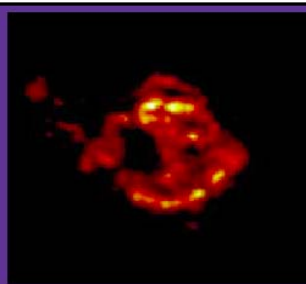
Etc.

Share all these images in Virtual Observatories

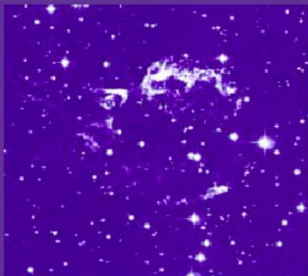
In order to COMPARE images you need MORE than the Web multi-λ views of a Supernova Remnant



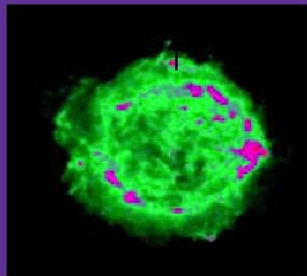
Shocks seen in the X-ray



Dust seen in the IR



Heavy elements seen in the optical

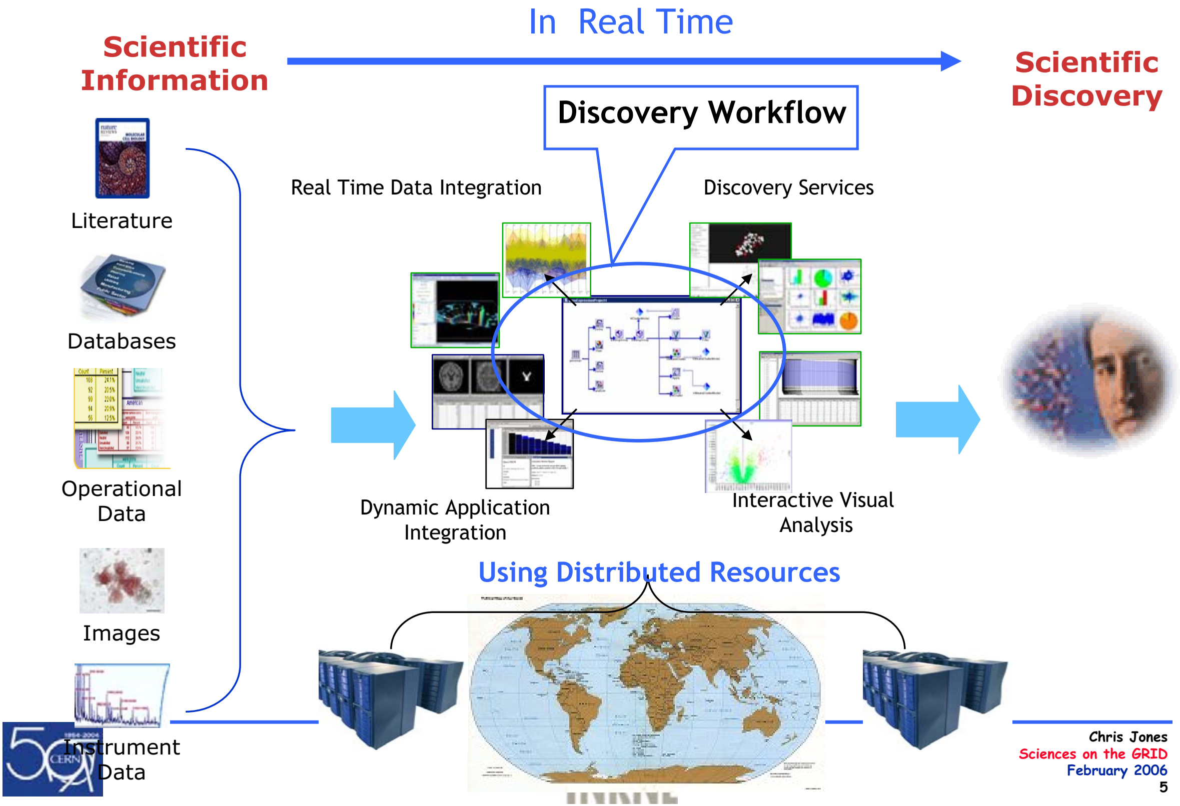


Relativistic electrons seen in the radio

Information / Knowledge GRIDS in Life Sciences

Many Active Projects

Discovery Net Snapshot



MyGrid

Personalised
extensible environments for
data-intensive
in silico experiments

in biology

Professor Carole Goble,
University of Manchester

Dr Alan Robinson,
EBI

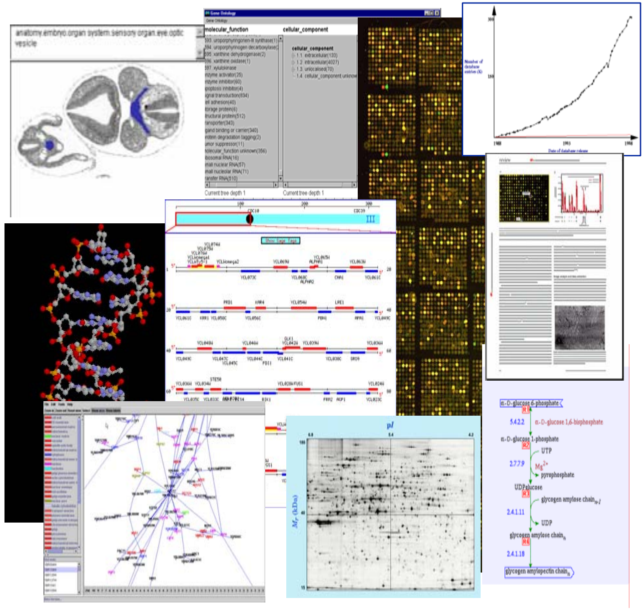
myGrid Project

Imminent 'deluge' of data

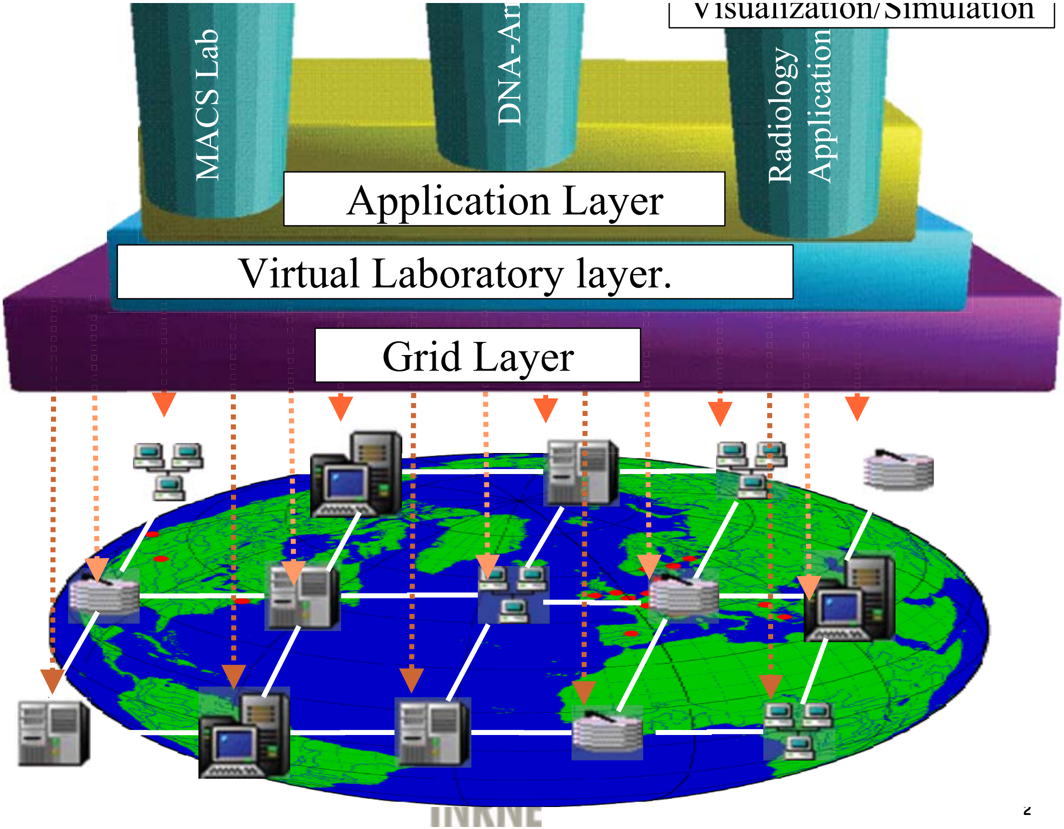
Highly heterogeneous

Highly complex and inter-related

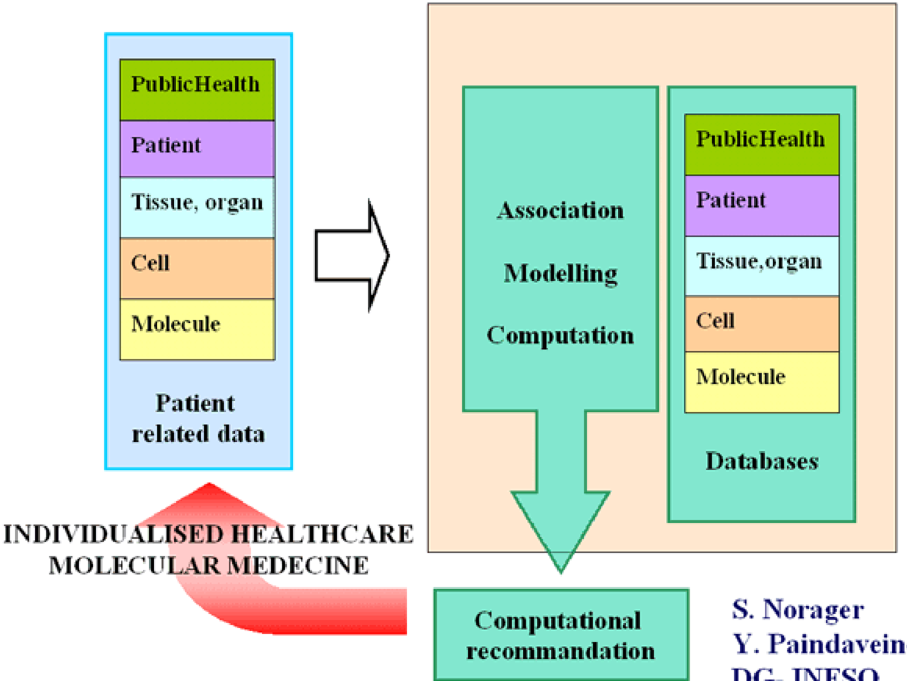
Convergence of data and literature archives



Professor Bob Hertzberger, University of Amsterdam, www.science.uva.nl



HealthGRID - DG-INFSO



S. Norager
Y. Paindaveine
DG- INFSO

Health GRIDS

HEALTHGRID

Second European HealthGRID Conference

Third conference in Oxford, March 2005

Definition of "Health" is broad (see next)

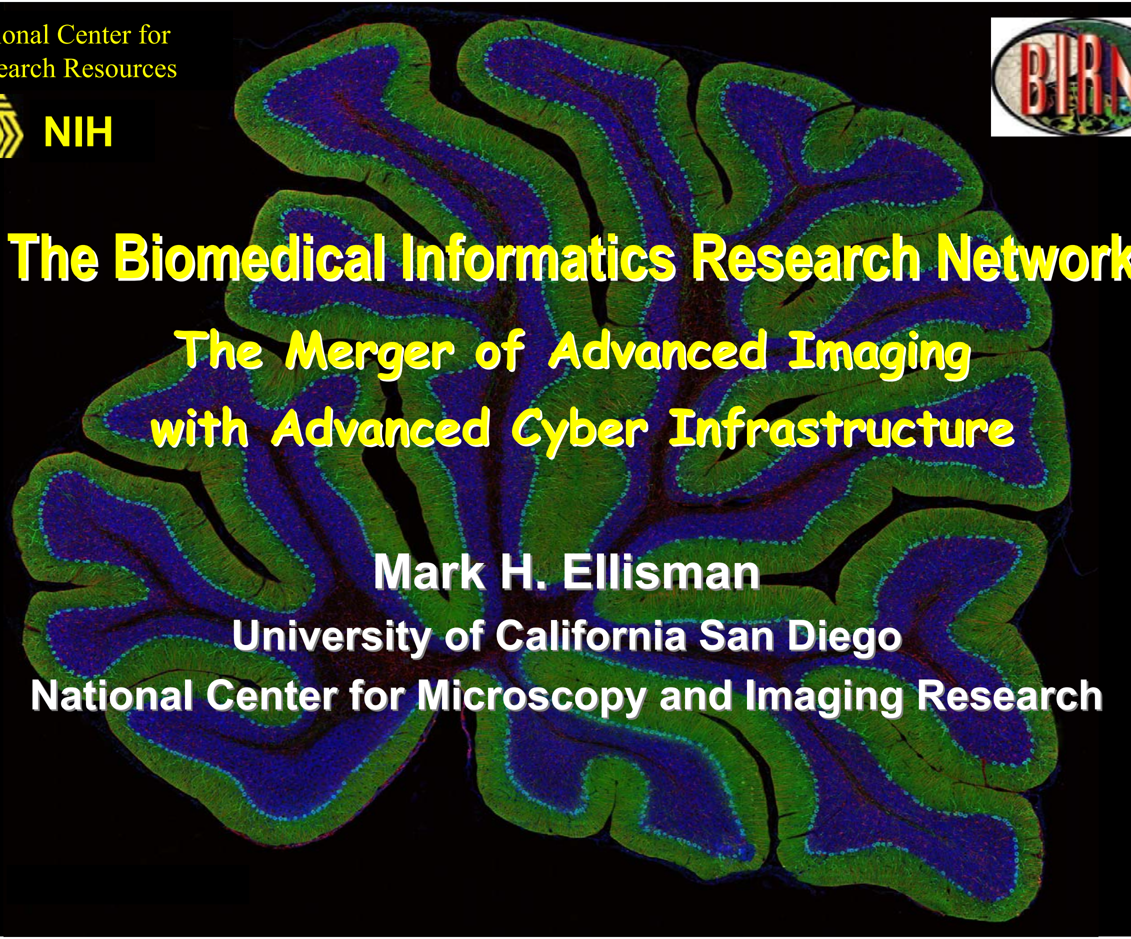
There is a useful Web Site with much information linked
http://www.healthgrid.org
Presentation from Ignacio Blanquer.pdf

White Paper for EU in final draft

There are many issues in the use of informatics in the domain of health, but there are many people working

National Center for
Research Resources

NIH



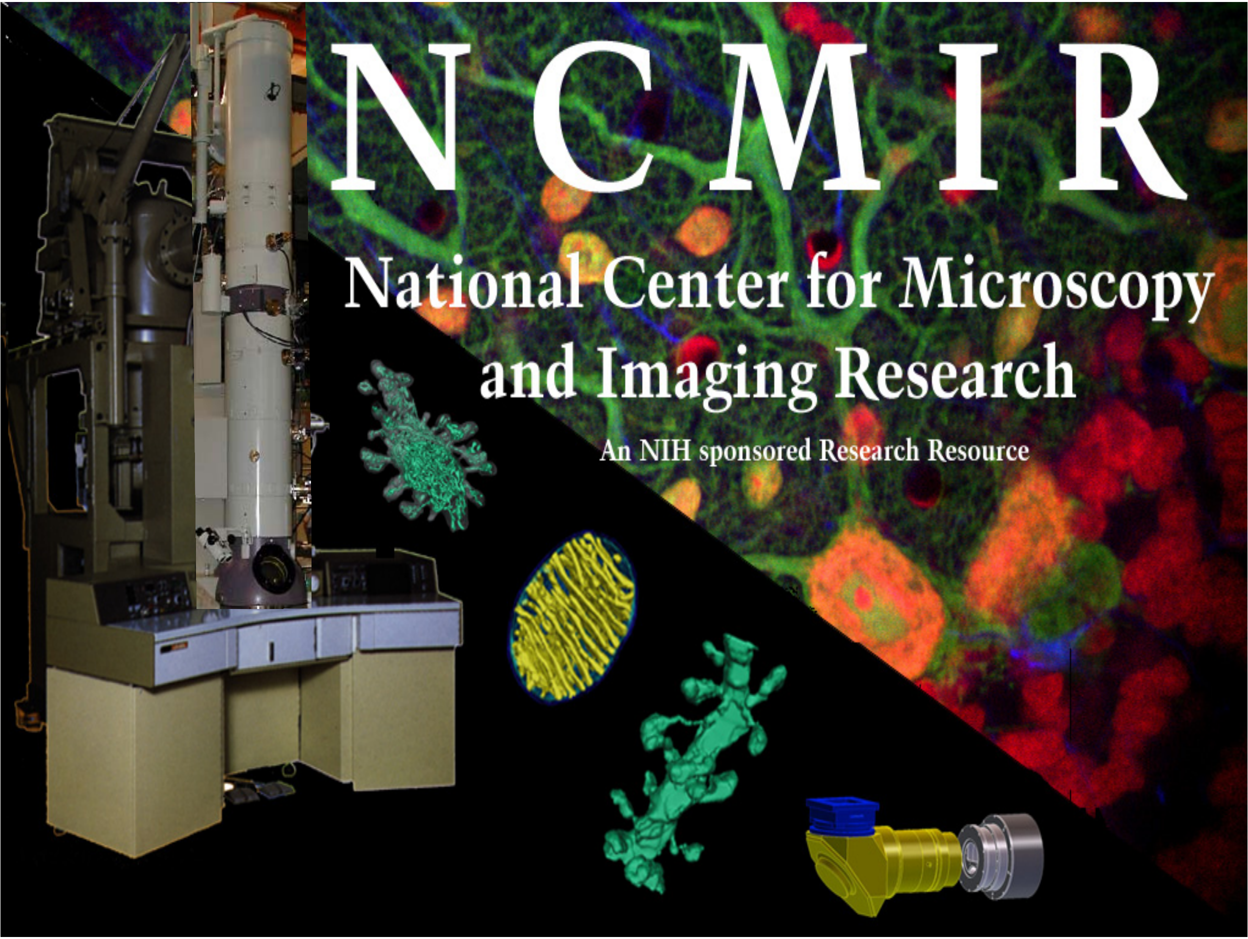
The Biomedical Informatics Research Network

The Merger of Advanced Imaging with Advanced Cyber Infrastructure

Mark H. Ellisman

University of California San Diego

National Center for Microscopy and Imaging Research


NCMIR
National Center for Microscopy
and Imaging Research
An NIH sponsored Research Resource

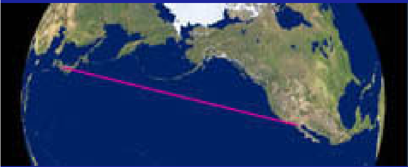
Modern 3D Transmission Electron Microscope

extreme penetration!

Ultra High Voltage EM @ Osaka Univ.

- **3 Million Electron Volts**
- **15 Meters Tall**
- **140 Tons**
- **$$$ > 50M US Dollars**
- **only one of these**

Trans-Pacific Telemicroscopy



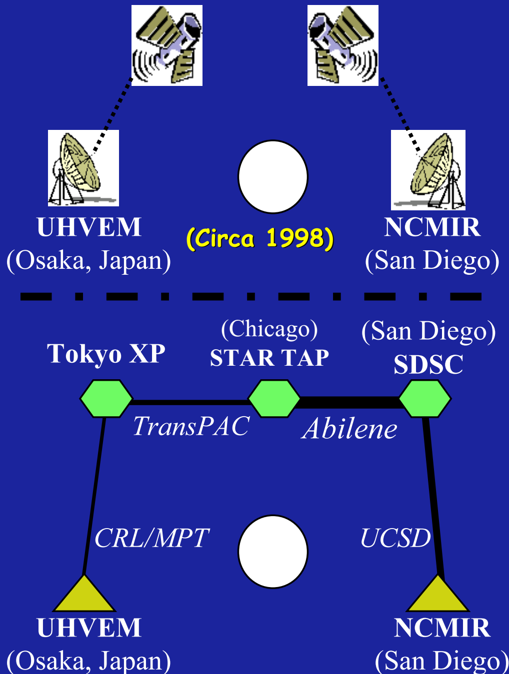
NET NEWS

Microscopy Across an Ocean

A big push by biologists to use computer networks to operate rare instruments from afar passed a major milestone on 25 June: Scientists took a spin on the world's most powerful electron microscope in Japan—while sitting 6000 kilometers away in California.

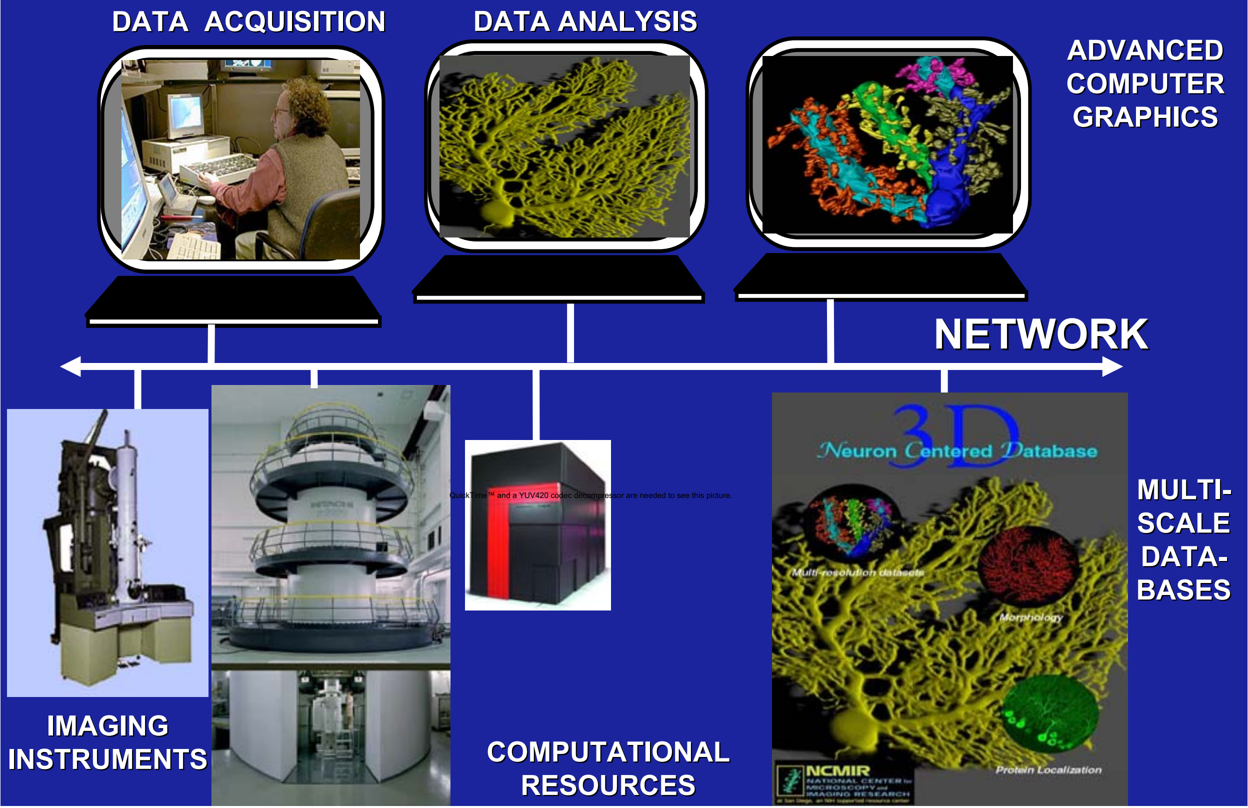
Six years ago, University of California, San Diego, neuroscientist Mark Ellisman thrilled audiences at a conference in Chicago by using the Internet to control an electron microscope in San Diego. Several U.S. agencies jumped in to fund projects for operating microscopes by remote control, and by now at least a dozen groups are doing so in the United States. Ellisman's team has since moved on to the Mount Everest of microscopes: Osaka University's Ultra High Voltage Electron Microscope, a 3,000,000-volt behemoth that can create three-dimensional images from much thicker samples (such as biological cells) than ordinary microscopes can. Ellisman and his U.S. and Japanese colleagues wondered if they could operate this instrument's roomful of controls from across the Pacific Ocean.

They showed they could. Over 5 hours, San Diego scientists imaged nerve cells from a rat and a frog without setting foot in Japan, controlling things like focus and specimen position across a private data line while the images came in across a satellite video link. Ellisman says this lays the groundwork for researchers all over the United States and Japan to borrow each others' specialized microscopes, probably via a high-speed Internet link, "within a year or two."



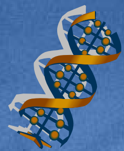
Now part of a production environment using IPv6

TELEMICROSCOPY & GRID - BASED COMPUTING
REMOTE ACCESS FOR DATA ACQUISITION AND ANALYSIS



International Infrastructures for Science and Industry

Earth Sciences



Life Sciences

Computer and Information Sciences

General Overview

Multidisciplinary Research

$E=MC^2$

Math and Physical Science

New Materials, Technologies and Processes

Social Sciences

Fabrizio Gagliardi

Director, Technical Computing, Microsoft EMEA

The e-Science Vision

- e-Science is about multidisciplinary science and the technologies to support such distributed, collaborative scientific research
  - Many areas of science are in danger of being overwhelmed by a 'data deluge' from new high-throughput devices, sensor networks, satellite surveys …
  - Areas such as bioinformatics, genomics, drug design, engineering, healthcare … require collaboration between different domain experts
- 'e-Science' is a shorthand for a set of technologies to support collaborative networked science

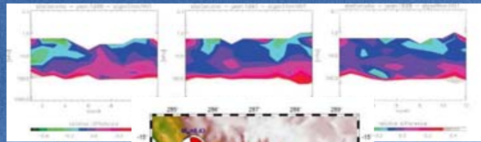
e-Science and the Grid

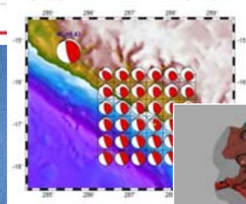
- Cyberinfrastructure and e-Infrastructure
  - In both the US and Europe there is a vision for the ‘cyberinfrastructure’ required to support the e-Science revolution
  - Set of Middleware Services supported on top of high bandwidth academic research networks
- Similar to vision of the Grid as a set of services that allows scientists – and industry – to <u>routinely</u> set up ‘Virtual Organizations’ for their research – or business
  - Many companies emphasize computing cycle aspect of Grids
  - The ‘Microsoft Grid’ vision should be more about data management than about compute clusters

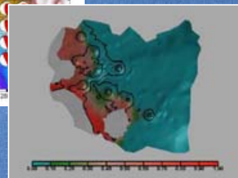
Need for e-Infrastructures

- Science, industry and commerce are more and more digital, process vast amounts of data and need massive computing power
- We live in a "flat" world:
  - Science is more and more an international collaboration and often requires a multidisciplinary approach
- Need to use technology for the good cause
  - Fight Digital/Divide
- Industrial uptake has become essential

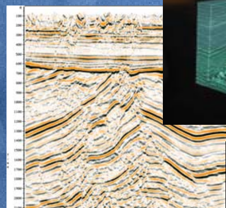
Some examples (from EU-EGEE): Earth sciences applications

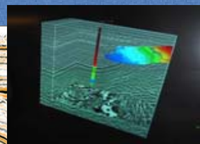


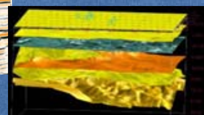




- **Satellite Observations:**
  - ozone profiles
- **Solid Earth Physics**
  - Fast Determination of mechanisms of important earthquakes
- **Hydrology**
  - Management of water resources in Mediterranean area (SWIMED)
- **Geology**
  - Geocluster: R&D initiative of the Compagnie Générale de Géophysique

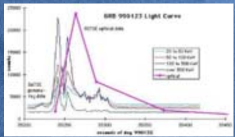


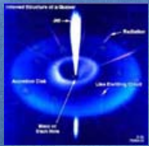


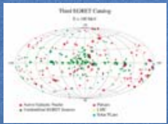


More examples: MAGIC, cosmic physics

- Ground based Air Cerenkov Telescope 17 m diameter
- Physics Goals:
  - Origin of VHE Gamma rays
  - Active Galactic Nuclei
  - Supernova Remnants
  - Unidentified EGRET sources
  - Gamma Ray Burst
- MAGIC II will come 2007
- Grid added value
  - Enable "(e-)scientific" collaboration between partners
  - Enable the cooperation between different experiments
  - Enable the participation on Virtual Observatories

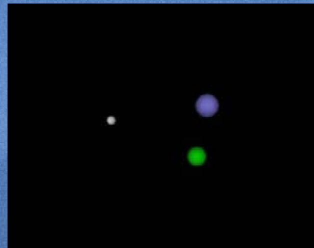






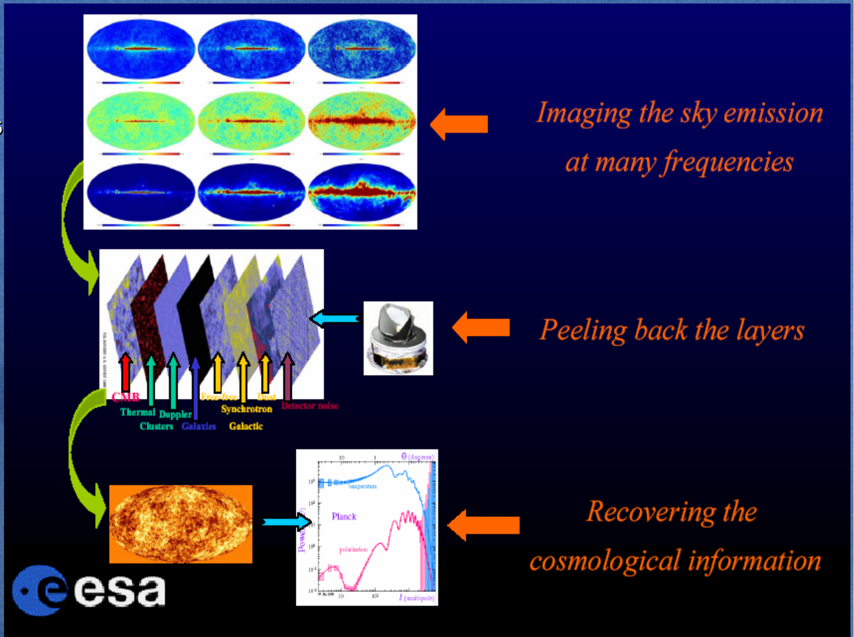
and more: Computational Chemistry

- The Grid Enabled Molecular Simulator (GEMS)
  - Motivation:
    - Modern computer simulations of biomolecular systems produce an abundance of data, which could be reused several times by different researchers.
      → data must be catalogued and searchable
  - GEMS database and toolkit:
    - autonomous storage resources
    - metadata specification
    - automatic storage allocation and replication policies
    - interface for distributed computation



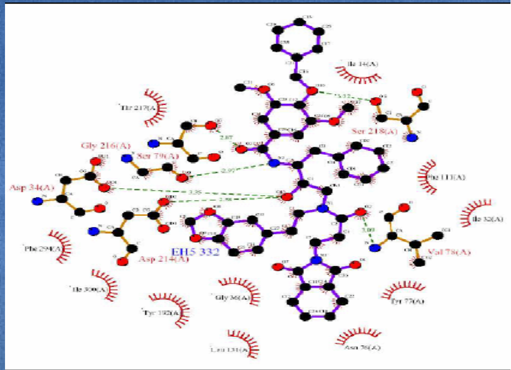
Planck: Cosmological studies

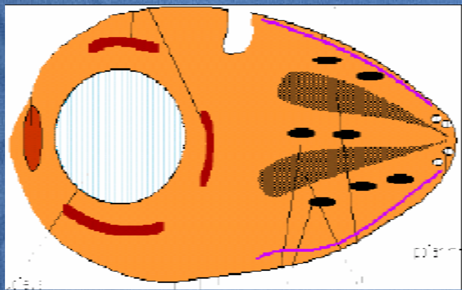
- On the Grid:
  - > 12 time faster
  - (but ~5% failures
- Complex data structure
  - → data handling important
- The Grid as
  - collaboration tool
  - common user-interface
  - Flexible environment
  - new approach to data and S/W sharing



Drug Discovery: a new hope for developing countries

- **Grid-enabled drug discovery process**
  - Reduce time required to develop drugs
  - Develop the next steps of the process (molecular dynamics)
- **Data challenge proposal for docking on malaria**
  - Never done on a large scale production infrastructure
  - Never done for a neglected disease
- **Data challenge during the summer**
  - 5 different structures of the most promising target
  - Output Data: 16,5 million results, ~10 TB
- **Added value**
  - Facilitates inclusion of developing countries
  - Tool to enhance collaboration building
  - Facilitate distributed software development for complex integrated workflows





Earth Science Activity

David Weissenbach (IPSL)
weissenb@ccr.jussieu.fr

EGEE is funded by the European Union under contract IST-2003-508833

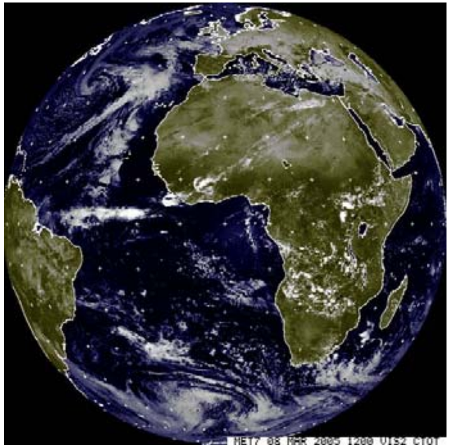
EARTH SCIENCE

The Earth is a complex system, centre of a large variety of phenomena.
Its description is separated in various independent domains with interfaces between them

Solid Earth Ocean Atmosphere

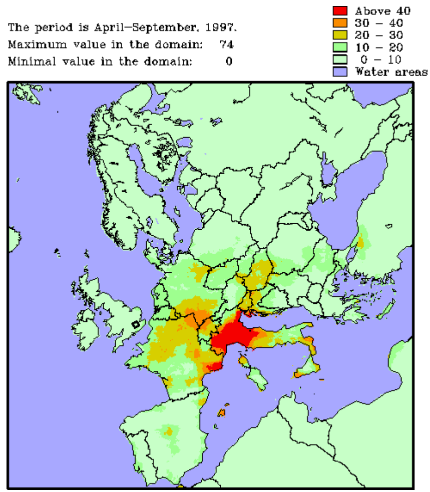
The approach could be based on physics, chemistry and/or biology

➔ The community is constituted by many small groups, that aggregate for projects (and separate afterwards).



Pollution (T.ostromsky ceco@parallel.bas.bg)

- BAS(Bulgaria): implementation of a large scale air pollution model on EGEE, for evaluation of the concentrations of a large variety of chemical species, responsible for the air pollution
- Model based on a parallel architecture
  - Distribution of computation
  - Development of new advanced splitting scheme (planned)



Difficulties:

- Large sets of data input and output: need a specific management
- Optimisation to take full advantage of the Grid

Recent past history

- Meta-computing and distributed computing early examples in the 80' and 90' (CASA, I-Way, Unicore, Condor etc.)
- EU-US workshop in Annapolis in 1999 on large scientific data bases: http://www.cacr.caltech.edu/euus/
- EU FP5 and US Trillium and national Grids
- EU FP6, US OSG, NAREGI/Japan…

GRIDs – IST in FP5 projects (~36m Euro)An integrated approach

Current situation: accomplishments and challenges

- Many Grids around the world, very few maintained as a persistent infrastructure (a good example is the Google Grid)
- Need for public and open Grids (OSG, EGEE and related projects, NAREGI, and TERAGRID, DEISA good prototypes)
- Persistence, support, sustainability, long term funding, easy access are the major challenges

Projects in Europe (I)

- Access to IT-resources (connectivity, computing, data, instrumentation…) for scientists:
  - **Providing e-Infrastructure**
    - **Géant2**
    - **EGEE**
    - **DEISA**
    - **SEE-GRID**
  - **Benefiting from e-Infrastructure**
    - **DILIGENT**
    - **SIMDAT**
    - **GRIDCC**
    - **CoreGRID**
    - **GridLab**
  - **Concertation: GRIDSTART, GridCoord**
  - **Grid mobility: Akogrimo**

Projects in Europe (II)

- **Sample of National Grid projects:**
  - Austrian Grid Initiative
  - DutchGrid
  - France:
    - e-Toile
    - ACI Grid
  - Germany
    - D-Grid
    - Unicore
  - Grid Ireland
  - Italy
    - INFNGrid
    - GRID.IT
  - NorduGrid
  - UK e-Science
    - National Grid Service
    - OMII
    - GridPP project

- EGEE

The EGEE project

- **Objectives**
  - consistent, robust and secure service grid infrastructure for many applications
  - improving and maintaining the middleware
  - attracting new resources and users from industry as well as science

- **Structure**
  - 71 leading institutions in 27 countries, federated in regional Grids
  - leveraging national and regional grid activities worldwide
  - funded by the EU with ~32 M Euros for first 2 years starting 1st April 2004

The EGEE project

- **Objectives**
  - Large-scale, production-quality infrastructure for e-Science
    - leveraging national and regional grid activities worldwide
    - consistent, robust and secure
  - improving and maintaining the middleware
  - attracting new resources and users from industry as well as science

- **EGEE**
  - 1st April 2004 – 31 March 2006
  - 71 leading institutions in 27 countries, federated in regional Grids

- **EGEE-II**
  - Proposed start 1 April 2006 (for 2 years)
  - Expanded consortium
    - > 90 partners in 32 countries (also non-European partners)
    - Related projects, incl.
      - *BalticGrid*
      - *SEE-GRID*
      - *EUMedGrid*
      - *EUChinaGrid*
      - *EELA*

Achievements

- **Infrastructure**
  - >170 sites
  - >15 000 CPUs
  - >5 PB storage
  - >10 000 concurrent jobs
  - >60 Virtual Organisations

- **Middleware**
  - Now at gLite release 1.4
    - Focus on basic services, easy installation and management
    - Industry friendly open source license

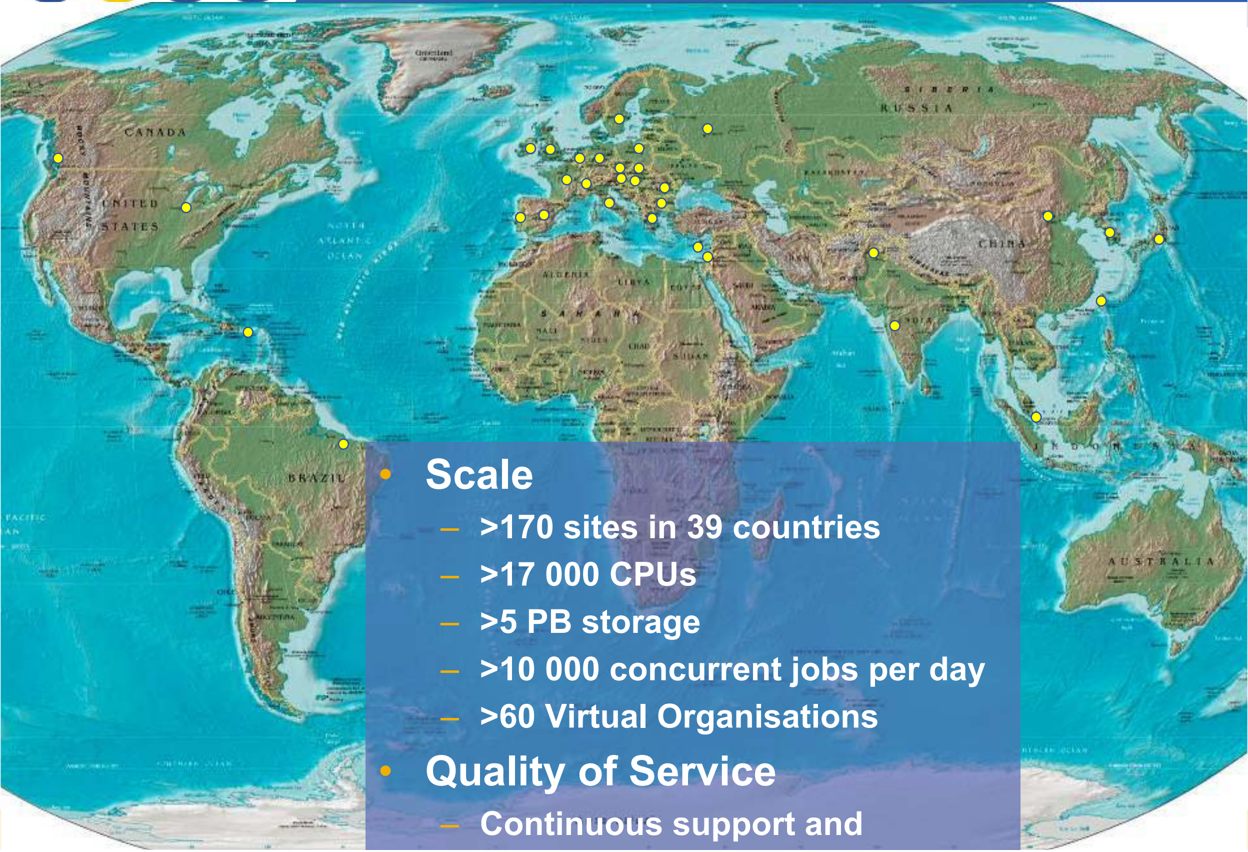
- **Applications: >20 applications from 7 doma**
  - High Energy Physics
  - Biomedicine
  - Earth Sciences
  - Astronomy
  - Geo-Physics
  - Financial Simulation

Another 8 applications from 4 domains are in evaluation stage

http://goc.grid.sinica.edu.tw/gstat/

- **Infrastructure**

EGEE Infrastructure



- **Scale**
  - >170 sites in 39 countries
  - >17 000 CPUs
  - >5 PB storage
  - >10 000 concurrent jobs per day
  - >60 Virtual Organisations
- **Quality of Service**
  - Continuous support and

EGEE services

- **Production service**
  - Based on the LCG-2 service
  - With new resource centres and new applications encouraged to participate
  - Stable, well-supported infrastructure, running only well-tested and reliable middleware

- **Pre-production service**
  - Run in parallel with the production service (restricted nr of sites)
  - First deployment of new versions of the middleware
  - Applications test-bed

- **GILDA testbed**
  - https://gilda.ct.infn.it/testbed.html
  - Complete suite of Grid elements and applications
    - Testbed, CA, VO, monitoring
  - Everyone can register and use GILDA for training and testing

Production Service

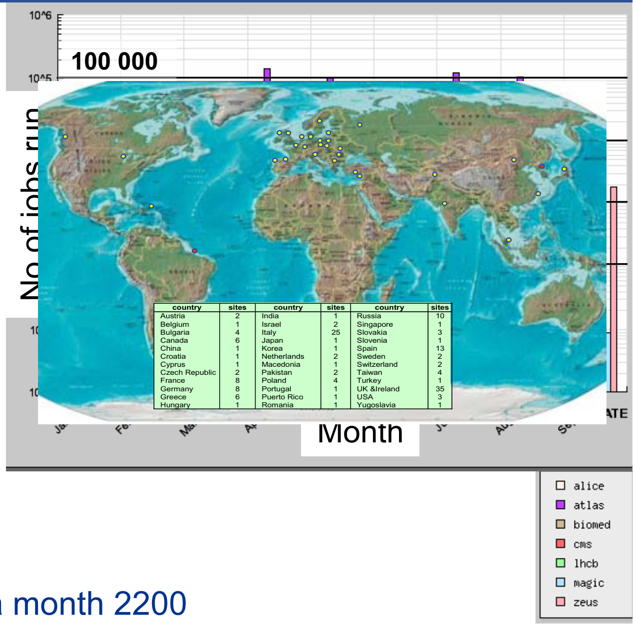
- **Scale of the production service**
  - ~16K CPUs/170 sites
  - Other national & regional grids: ~60 sites, ~6000 processors

  → greatly exceeds no of sites planned for the end of EGEE

- **Interoperability demonstrated with OSG (ongoing work with ARC)**

- **> 2.2 million jobs Jan-Oct 2005**
  - Daily averages, sustained over a month 2200 – 10 100

→ ~6 M kSI2K.cpu.hours ≅ ~700 cpu years

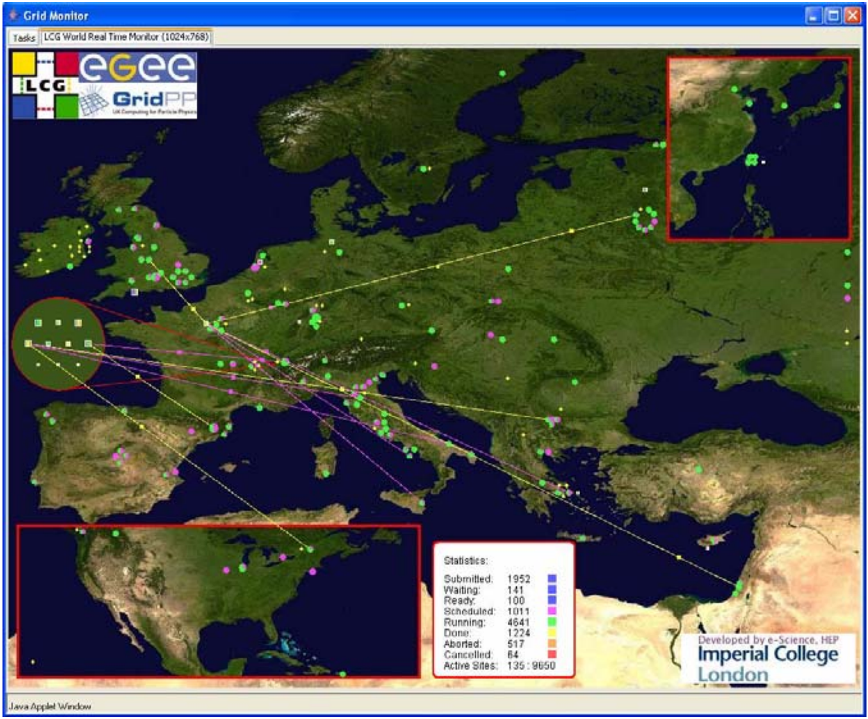


country	sites	country	sites	country	sites
Austria	2	India	1	Russia	10
Belgium	1	Israel	2	Singapore	1
Bulgaria	4	Italy	25	Slovakia	3
Canada	6	Japan	1	Slovenia	1
China	1	Korea	1	Spain	13
Croatia	1	Netherlands	2	Sweden	2
Cyprus	1	Macedonia	1	Switzerland	2
Czech Republic	2	Pakistan	2	Taiwan	4
France	8	Poland	4	Turkey	1
Germany	8	Portugal	1	UK &Ireland	35
Greece	6	Puerto Rico	1	USA	3
Hungary	1	Romania	1	Yugoslavia	1

Production Service

LCG2 Real Time Monitor

- Java tool
- Displays jobs running (submitted through RBs)
- Shows jobs moving around world map in real time, along with changes in status



http://gridportal.hep.ph.ic.ac.uk/rtm/

(snapshot 18 November 2005)

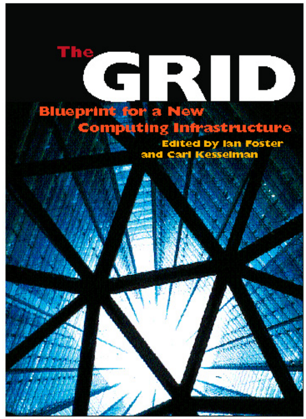
Books, Details…

The Grid - A Book

"Blueprint for a New Computing Infrastructure"

A Serious Textbook - university course level

- Ian Foster and Carl Kesselman are the editors
- Two chapters of Overview
- Twenty chapters of detailed areas written by appropriate experts:
  - challenges to be met
  - review of the state of the art
- Morgen Kaufmann publishers
- http://www.mkp.com/



"This is a source book for the history of the future."

Vint Cerf, Senior Vice President,
Internet Architecture and Engineering,
MCI Communications

And another and newer book…

Grid Computing

Making the Global Infrastructure a Reality

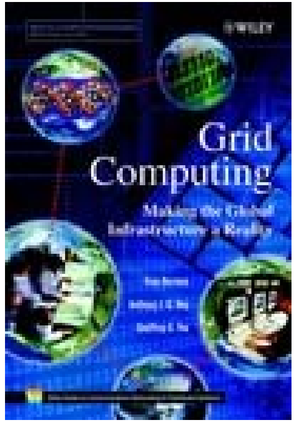
Fran Berman

Geoffrey C. Fox

Anthony J.G. Hey

Wiley

www.wiley.com



Global GRID Forum

First Meeting, joining separate Grid Forum initiatives, Amsterdam in April 2001, sold out with 350 people

Subsequent meetings in Washington, Rome, Toronto, Edinburgh, Chicago, excellent attendance

And growing industry attendance

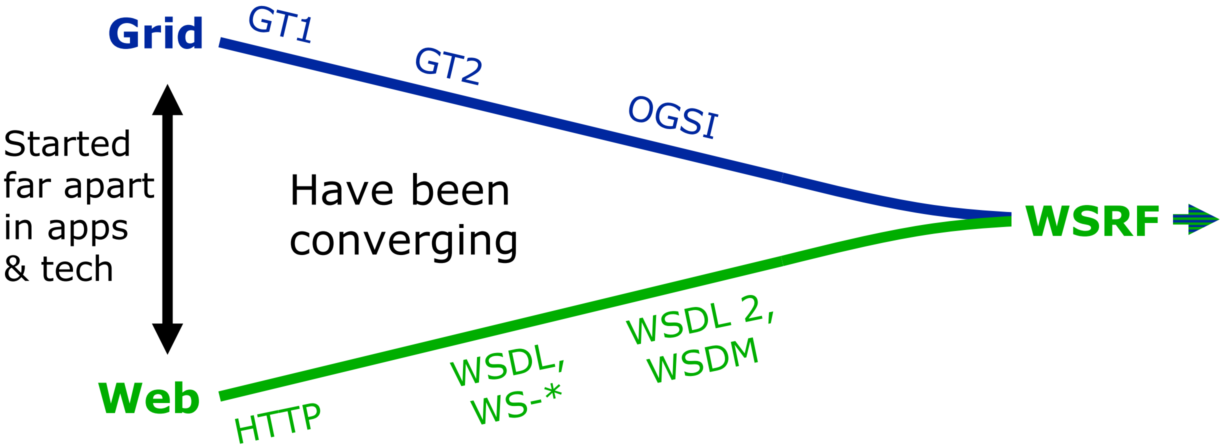
As the role that industry will play in this open source - open interface unfolds

Seventh Meeting, March 2003, Japan

Eight Meeting, June 2003, Seattle…

www.ggf.org

Grid and Web Services: Convergence: Yes!

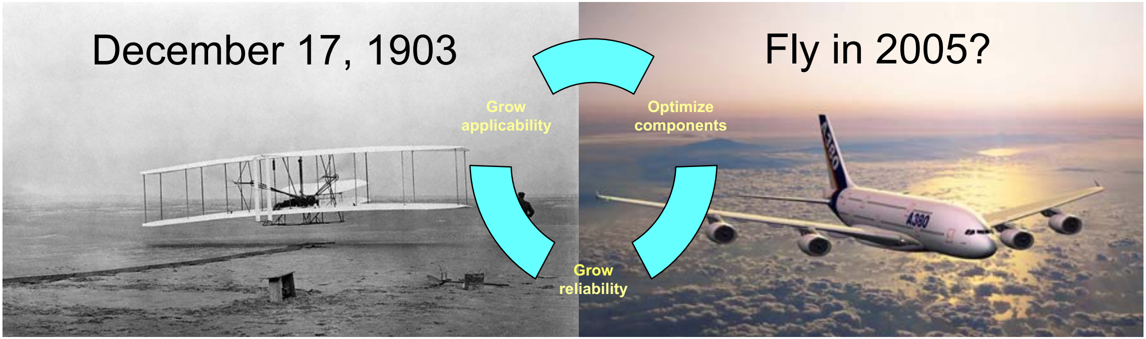


The definition of WSRF means that Grid and Web communities can move forward on a common base

So where are we?
When is this all going to happen?

The „Kitty Hawk Project": will it fly?

Let's make it fly!

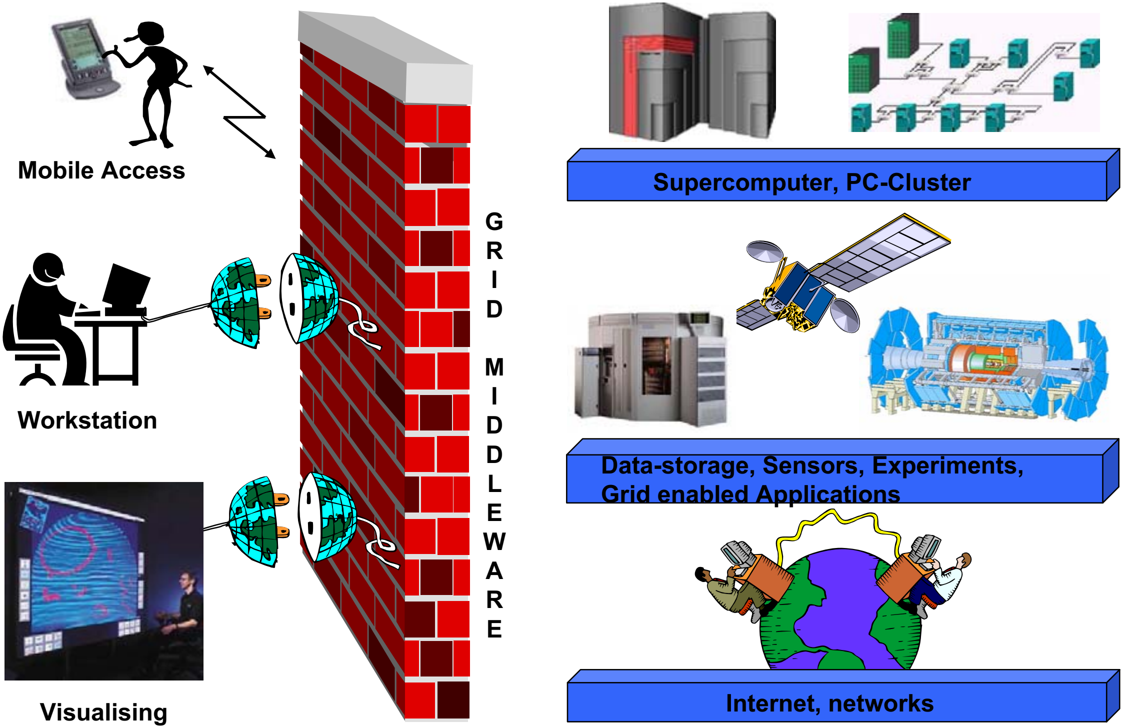


New concepts

100 years of optimization and improvements

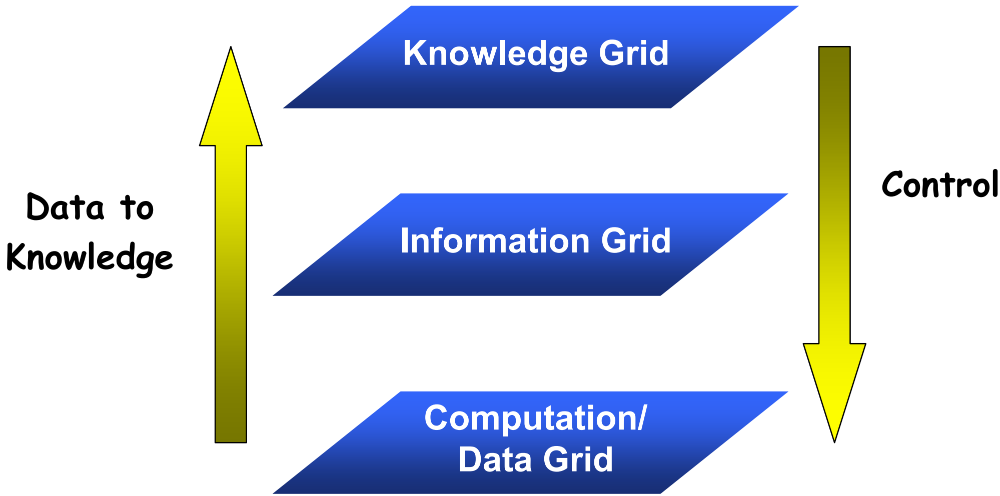
Thank You

The "One-Stop Shopping" view of the GRID



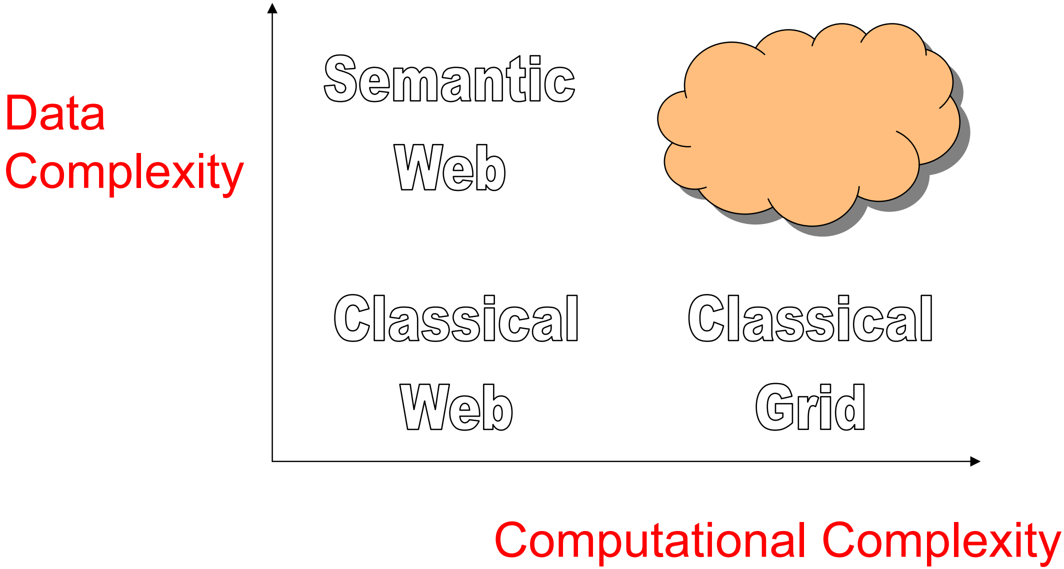
Hoffmann, Putzer, Reinefeld

Three Layer GRID Abstraction



•Data	Un-interpreted bits and bytes
•Information	Data equipped with meaning
•Knowledge	Information applied to achieve a goal, solve a problem or enact a decision

Databases in the Grid



A final word

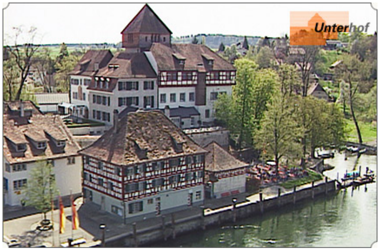
Of course this vision raises many questions and obstacles that must be faced at the same time as the technology is developed:

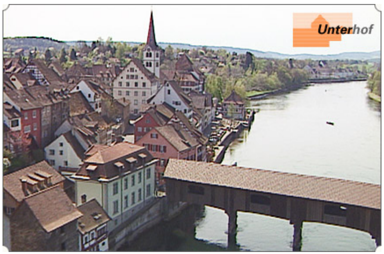
- questions of security,
- access rights,
- authentication,
- ethics,
- policies etc.

This is true of most new technologies

It is important that these issues are faced directly and not later as has been all too often the case.

PharmaGRID 2004





2004 PharmaGRID Retreat
July 7th-9th 2004

Seminarhotel Unterhof,

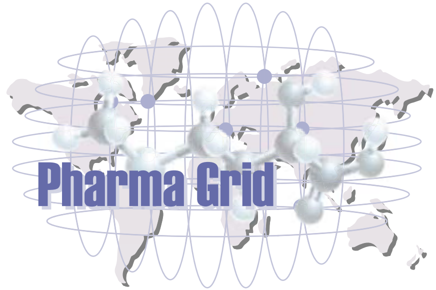
Diessenhofen, Switzerland

Aims to bring together industry professionals, stakeholders and leading GRID practitioners

For more information

www.pharmaGRID.com

Pharma Grids: Integration Potential in Both the Private and the Public Sectors



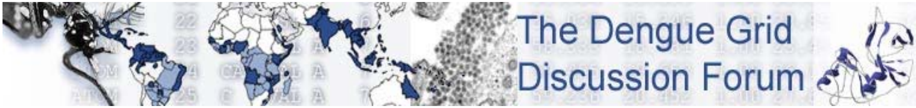
Disease knowledge resource:

- Third world diseases
- Rare diseases (orphan diseases)
- Lead discovery for third world and/or rare diseases

Educational resource:

- Training of health care professionals (eLearning) (in particular in remote locations)
- On-line assistance to health care professionals (in particular in remote locations)

The University of Basel Contribution to the Dengue Knowledge Grid



www.bc2.ch/grid-forum

March 18-19
2004

UNI BASEL

SIB

BIOZENTRUM

FMI

NOVARTIS

Ajilon

Further Initiatives towards Knowledge Resources for Neglected Diseases

NIH:

NCRR (Natl. Center for Research Resources) and

Office of Rare Diseases; contact Dr Howard Bilofsky, bilofsky@pcbi.upenn.edu

Virginia Bioinformatics Institute (Prof. B.Sobral)

https://research.vbi.vt.edu

PathPort: information resource on Anthrax, Plague, Smallpox, many types of hemorrhagic fevers

PharmaGrid 2004 (chris.jones@cern.ch), co-sponsored and co-organised by PRISM (www.prismforum.org)

There are issues at each layer

- Knowledge Discovery and Data Mining Services
- Ontologies / Knowledge Representation
- Workflows
- Integration of Applications
- Distributed Data Access / Information Retrieval
- Administration of Virtual Organisations
- Basic Grid Technology Layer

Data Acquisition

- In 1993 WTSI (then the Sanger Centre – named after double Nobel prize winner – Fred Sanger) started with just a few gigabytes of data

If WTSI storage had stayed in line with Moore's Law

1993	**94**	**1995**	**96**	**1997**	**98**	**1999**	**2000**	**2001**	**02**	**2003**	**04**	**2005**
20GB		**40GB**		**80GB**		**160GB**		**320GB**		**640GB**		**1.3TB**

- Within the first 6 years (1993-1999) our data store had grown to over 4.5 Terabytes (4,500 gigabytes).