

Developments in BioPharma Grid Computing

Rowan Gardner, BioLauncher Ltd



PharmaGRID 2004



60 Delegates, a beautiful location, good food and excellent speakers...



Speakers

- Prof. Denis Noble, Applications in Systems Biology, a vision of the future of GRID
- Denise Ecklund, PhD, Architectures to support data integration and federation
- Ken Buetow, PhD CaBIG, The Cancer BioMedical Informatics GRID
- John Wilbanks, Semantic Web
- Professor Mark Ellisman, BIRN, BioMedical Informatics Research Network
- Professor Andy Lawrence AstroGRID
- Prof. Bob Hertzberger Virtual Laboratories
- First Hand Experience of GRID in the Pharmaceutical Industry
 - Novartis, Eli Lilly, Johnson and Johnson,
- Prof. Carol Goble, Delivering on the Promise of GRID
- Prof. Manuel Peitsch, A Vision for PharmaGRID

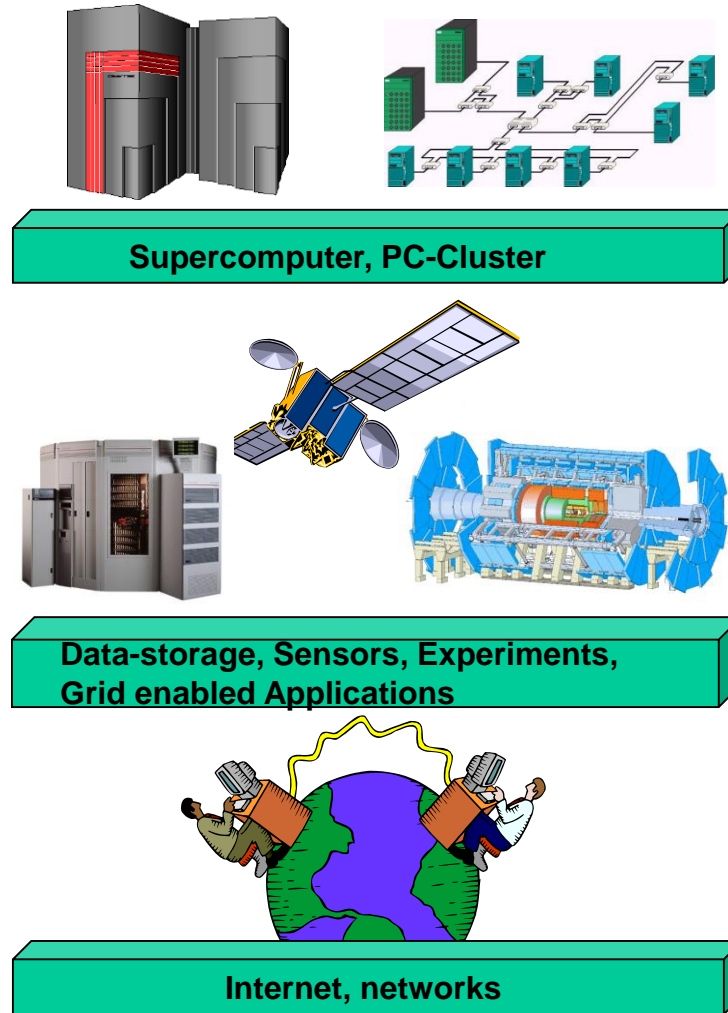
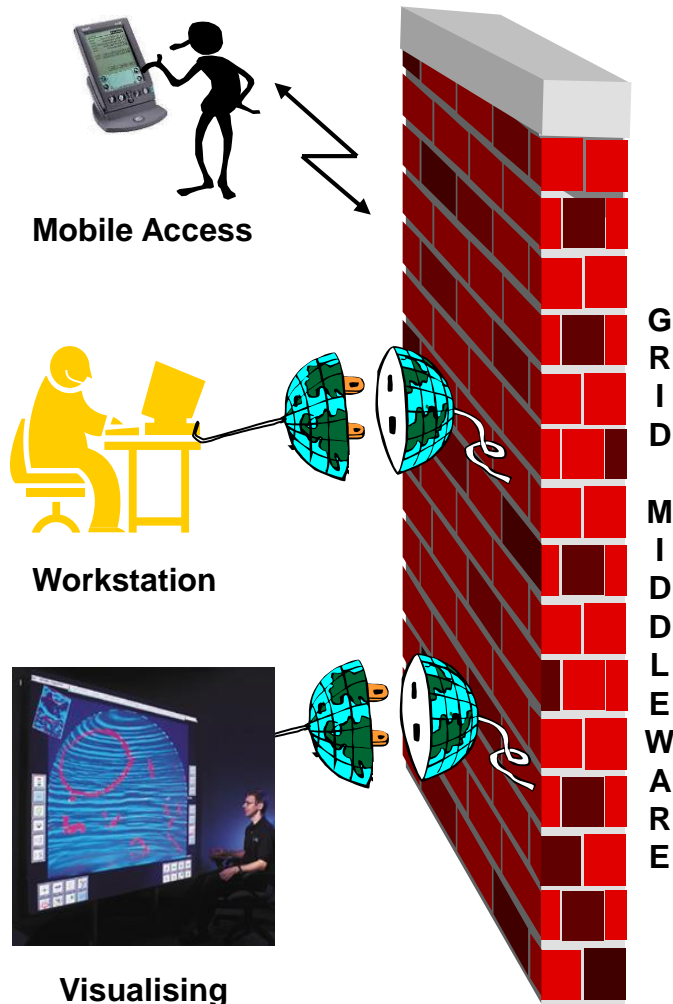


Today's talk

- A Vision for PharmaGRID
 - Why it is important
 - Evolving standards
- What has been achieved to date?
 - Examples from industry and academia
- Semantic technologies and the GRID
- The Way Forward



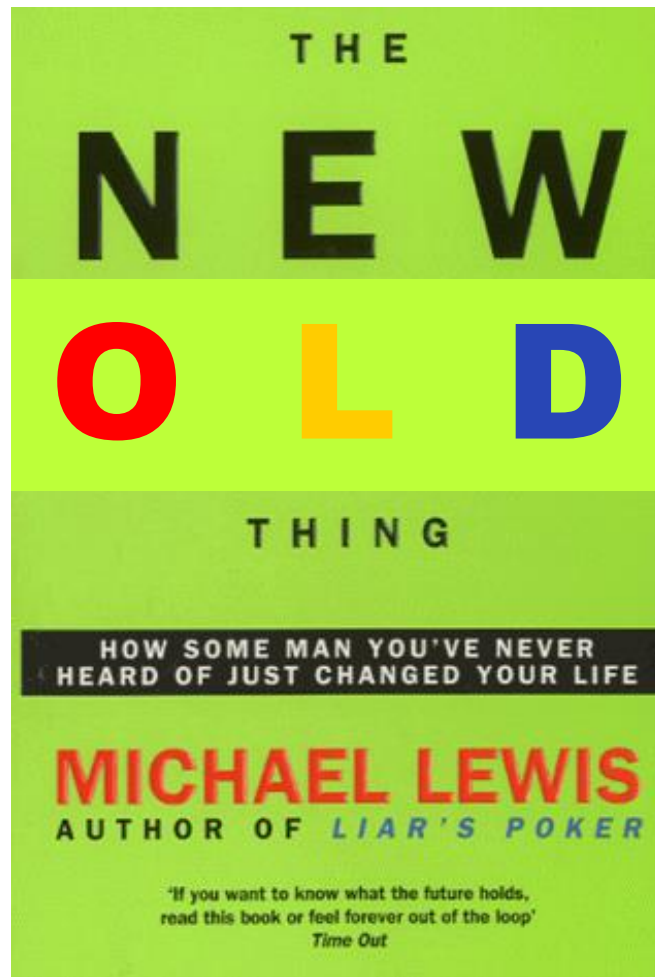
A Vision of GRID: One Stop Shopping



Hoffmann, Putzer, Reinefeld



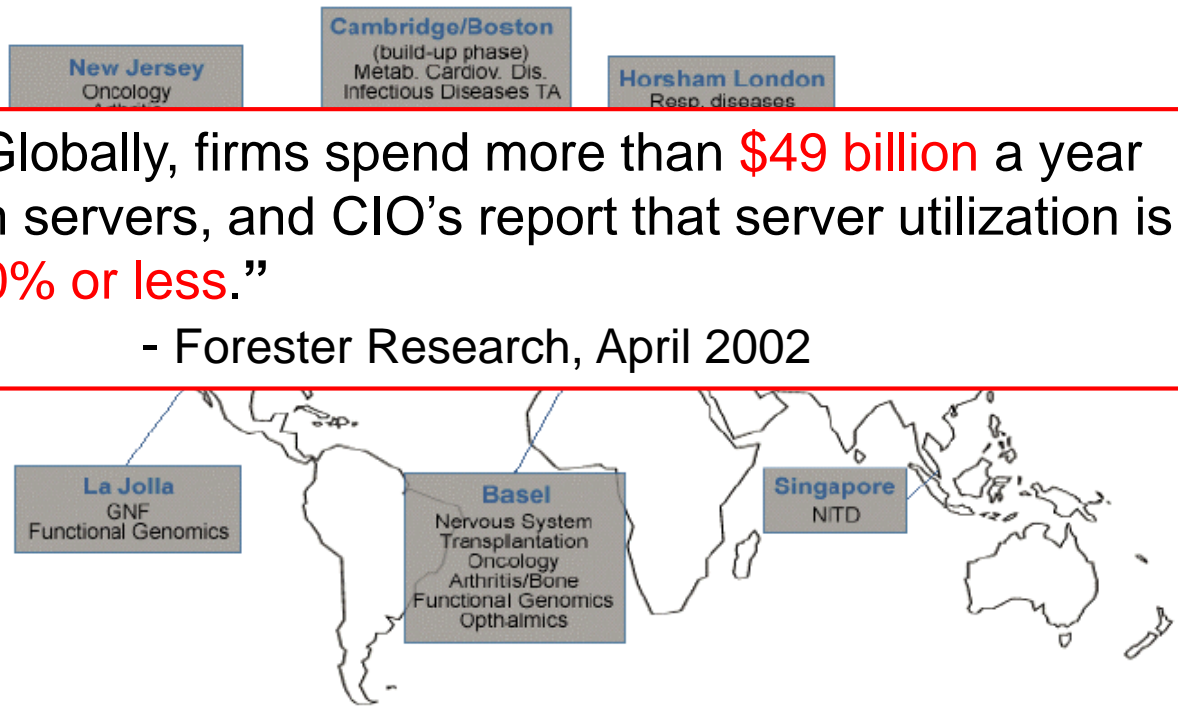
How Significant is GRID?



We Work in a Distributed World

Geographical complexity.

Novartis Pharmaceuticals Research Worldwide Community 2002
3000 Scientists

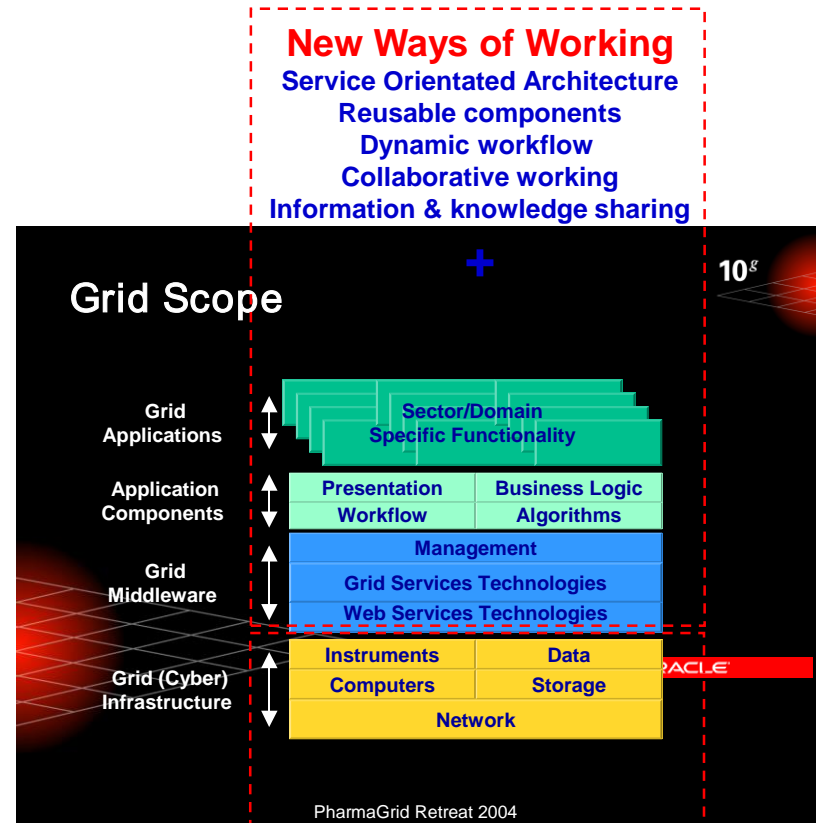
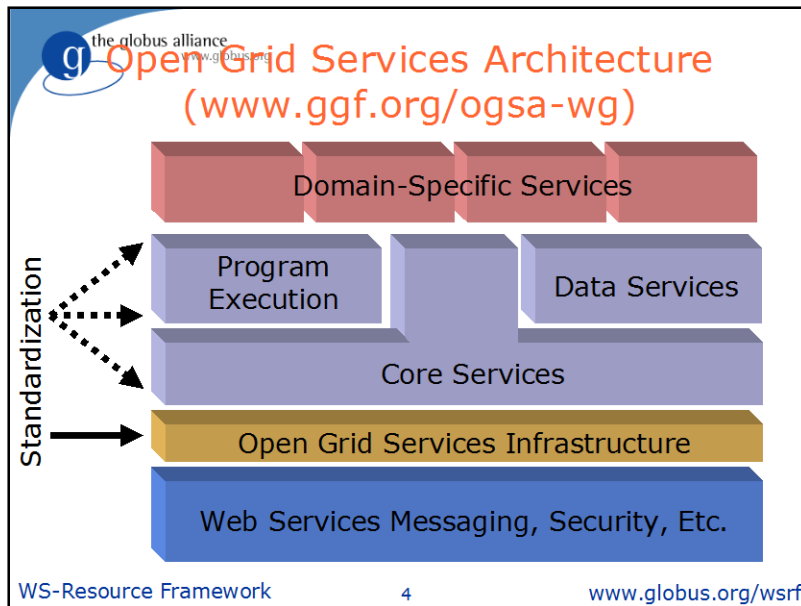


“Globally, firms spend more than **\$49 billion** a year on servers, and CIO’s report that server utilization is **60% or less.**”
- Forester Research, April 2002



The industry has to deliver more from its investment

GRID Architecture

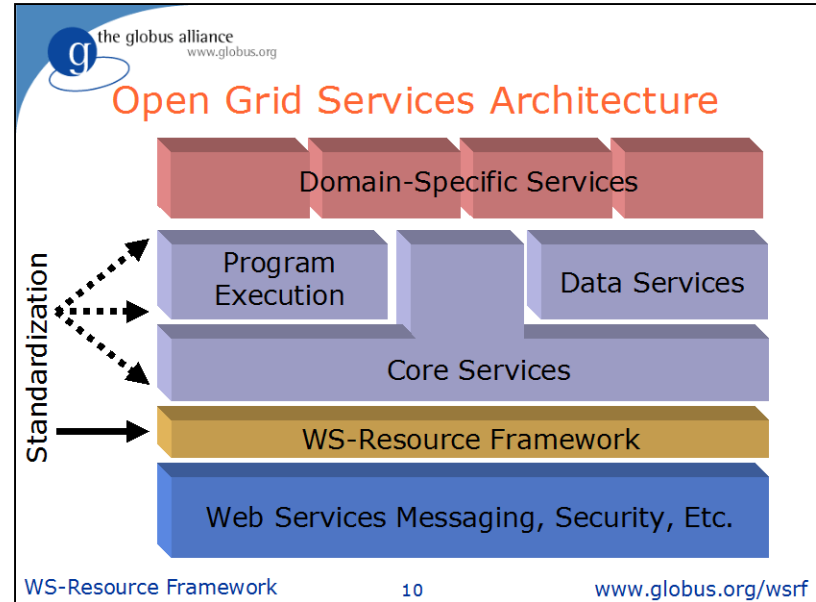
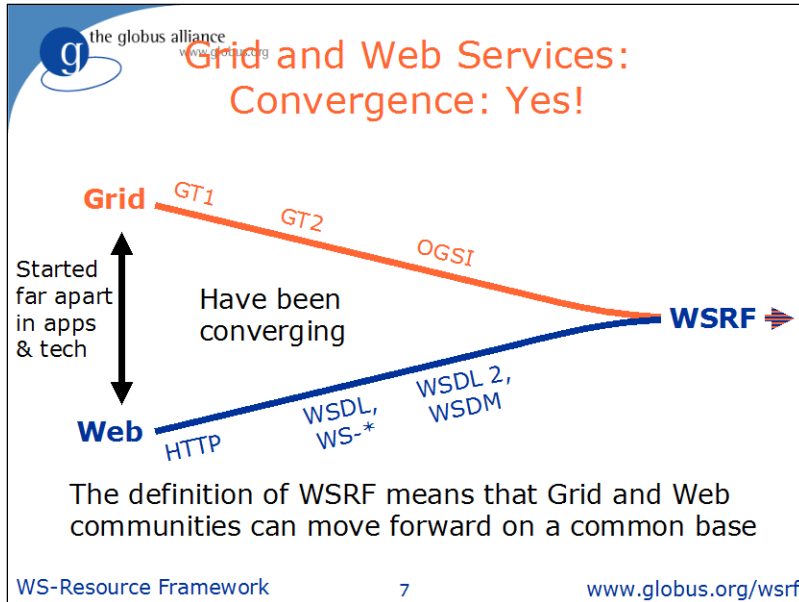


New Ways of Working
 Service Orientated Architecture
 Reusable components
 Dynamic workflow
 Collaborative working
 Information & knowledge sharing

Increased ROA
 Infrastructure resource sharing
 Automated monitoring & management



Standards Convergence



Convergence of GRID and Web Services;
So if you developing using Web services you get GRID for free...



What has been achieved?

- “Crunch-GRIDs” HPC resources typically PC clusters
 - Cycle scavenging grids
- “Knowledge Grids”
 - Portal style access to a set of federated databases and applications and experimental resources
- Validation of GRIDs in the clinical domain
- e-Collaboration applications



Crunch GRID: GRID.org

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

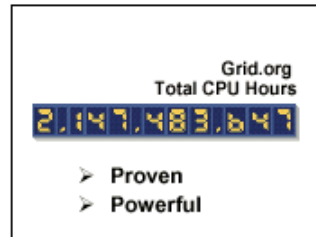
➤ Results

- 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

Grid.org is...



- 2.5M Nodes
- 191+ Countries



- Fully hosted, scalable
- In production for over 3 years

Intel said the software's security is "robust end to end". Barrett said he was going to allow it to be run inside his company.

- Secure
- Trusted by the public and the world's leading institutions

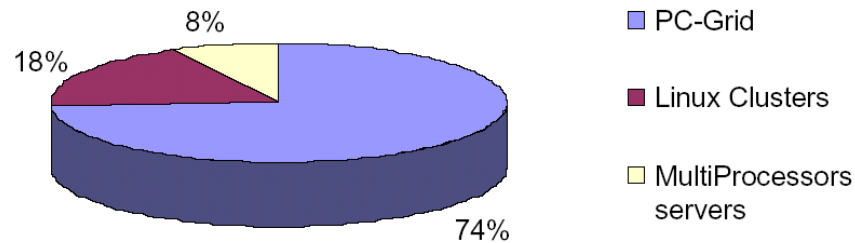


Crunch GRIDs: Novartis

HPC global environment

- Worldwide distribution of ~30 HPC systems.
- **Big numbers** ;-)
 - 1.8TB Ram.
 - 122TB Disk space usable.

HPC CPU cycles distribution



GRIDs in the Regulated Domain

FD : Simulations (Validated environment) *Jeff Mathers*

- Regulatory Compliance

- The key for grid success!

- Cross-Pharma collaboration formed

- Novartis, GSK, Pfizer, Merck, and J&J

- Produce “guidance” or “best practice” for how to ensure quality of data and systems management to allow for using the grid with validated applications.



Johnson & Johnson
PHARMACEUTICAL RESEARCH
& DEVELOPMENT
DIVISION OF JANSSEN PHARMACEUTICA N.V.



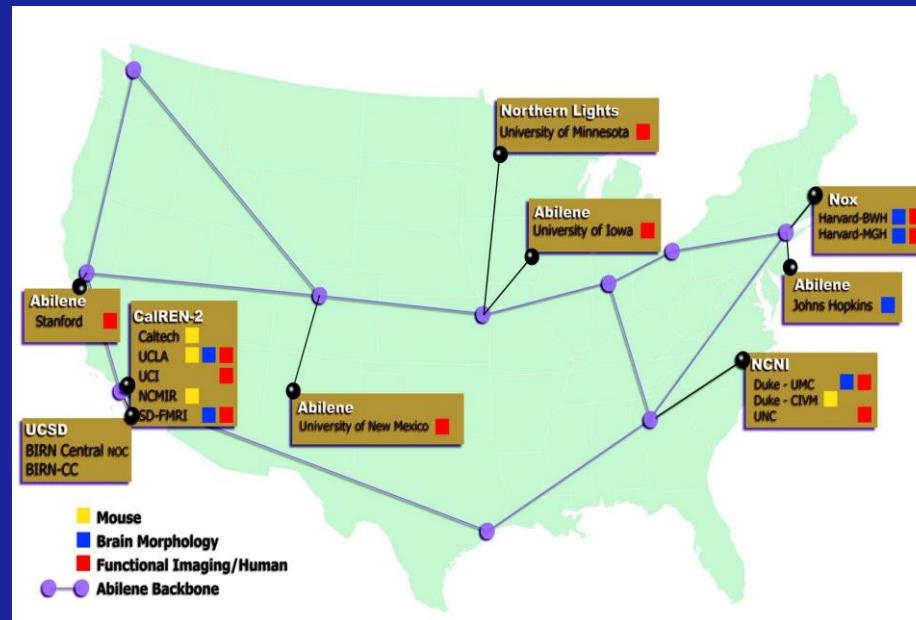
Knowledge Grids: BIRN

BIRN Network



- A stable, robust, shared network and distributed database environment across **>15 institutions**, tailored to the BIRN collaborations.
- **Extensible tools** and IT infrastructure that can be reused.
- Established **Cyber-infrastructure** for a **neuroscience data grid**
- Involves a large scale **data integration** effort
“DATA MEDIATION”
- Soon the NIH’s BIRN will begin to include areas of biomedical research other than Neuroscience

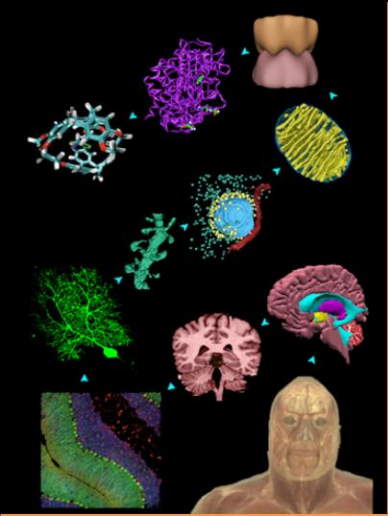
IT Infrastructure to hasten the derivation of new understanding and treatment of disease through use of distributed knowledge



**It will no longer matter where data, instruments and computational resources are located!*

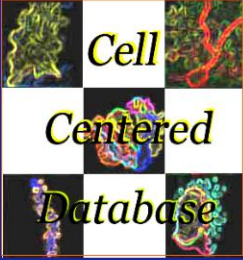
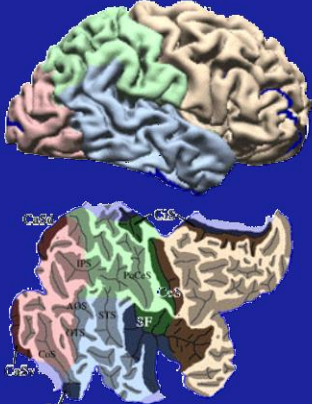
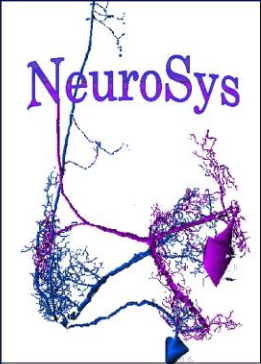
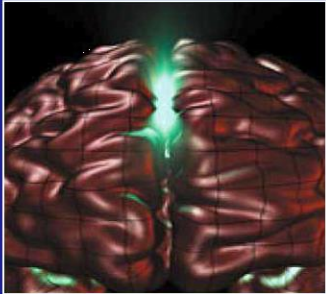


Federation of Data



Federation of Brain Data

- National Partnership for Advanced Computational Infrastructure (1996-04)
- Integrating brain data across scales and disciplines



UCSD
M. Martone, A. Gupta, M. Ellisman

UCLA
Art Toga

Montana State Univ
Gwen Jacobs

Washington Univ
David Van Essen

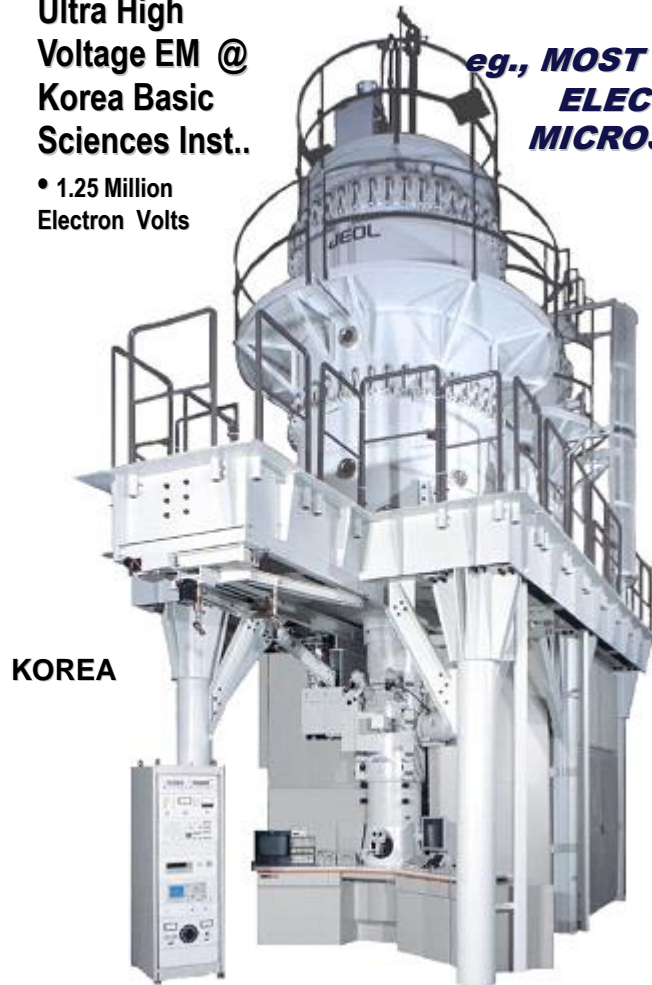


BIRN Instruments

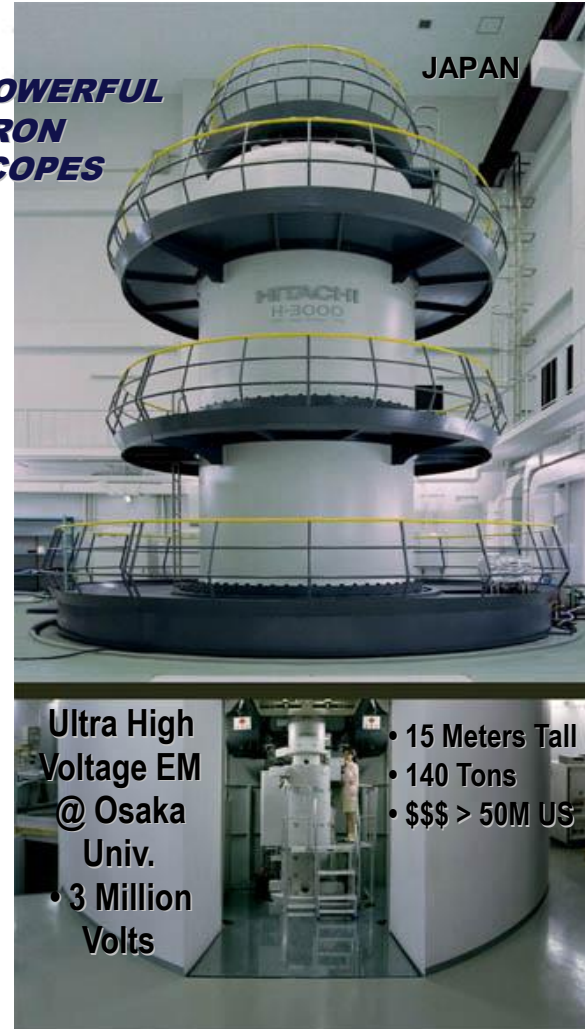
Access to Unique Instruments

Ultra High Voltage EM @ Korea Basic Sciences Inst..

- 1.25 Million Electron Volts



eg., MOST POWERFUL ELECTRON MICROSCOPES



Ultra High Voltage EM @ Osaka Univ.
• 3 Million Volts

- 15 Meters Tall
- 140 Tons
- \$\$\$ > 50M US



Semantic Web and Ontology

- 5 out of the first 6 talks at PharmaGRID 2004 highlighted ontology/other knowledge representations as important
- Recognition of the importance of ontologies to GRID
 - Retrieve relevant data from multiple resources
 - Annotation
 - e-Collaboration



Semantic Challenges

Why The Semantic Web Matters In Life Sciences R&D

- Scientific data and applications require rich semantic representation (both internal and external apps and data)
- Most life sciences organizations (large & small, commercial & academic) currently lack experience and exposure to new semantic technologies
- Over the next 5 years, software and IT professionals in the life sciences will experience unprecedented demand for semantically rich, data driven applications
- Semantic technologies and standards developed in the past 5 years must dramatically improve the effectiveness and efficiency of biopharmaceutical discovery

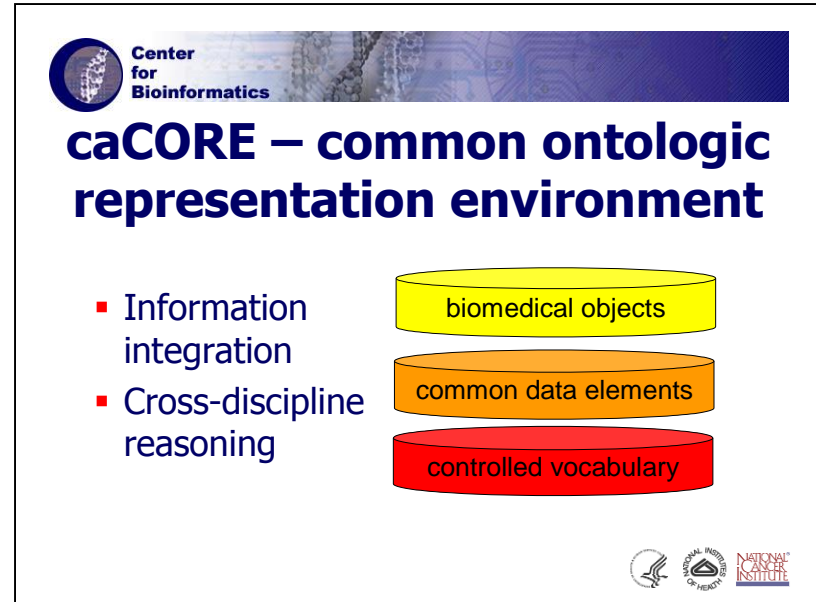
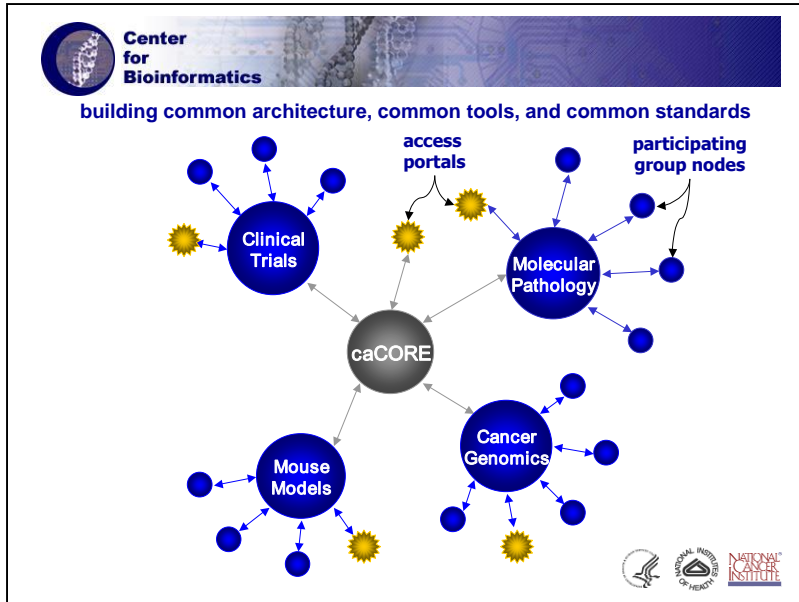
SW-LS and Grid: Business Case

- **Treat knowledge as a single, grid-enabled corporate asset**
 - More efficient knowledge use
 - “Aggregate” the knowledge – data and interpretation – into a single, extensible web, regardless of location
 - After aggregating, deploy context-based search
 - Long-term:
 - Knowledge-driven data stratification (annotation, ranking, *hypothesis generation / elimination*)
- **Maximize the value of knowledge through *automation* and *reuse* – tie to web services**

- Start thinking about how you will build knowledge representations
- Start exploring ontology browsers such as Haystack



CaBIG: Applying Ontologies



- **Goal:** A virtual web of interconnected **data, individuals, and organizations** redefines how research is conducted, care is provided, and patients/participants interact with the biomedical research enterprise



The “Kitty Hawk Project”: will it fly?

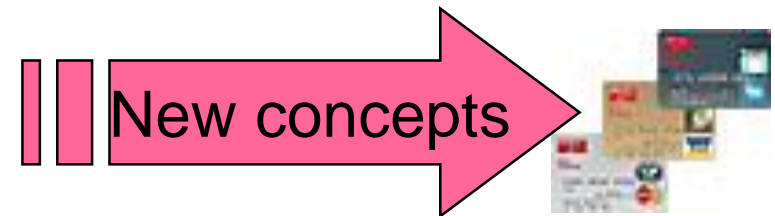
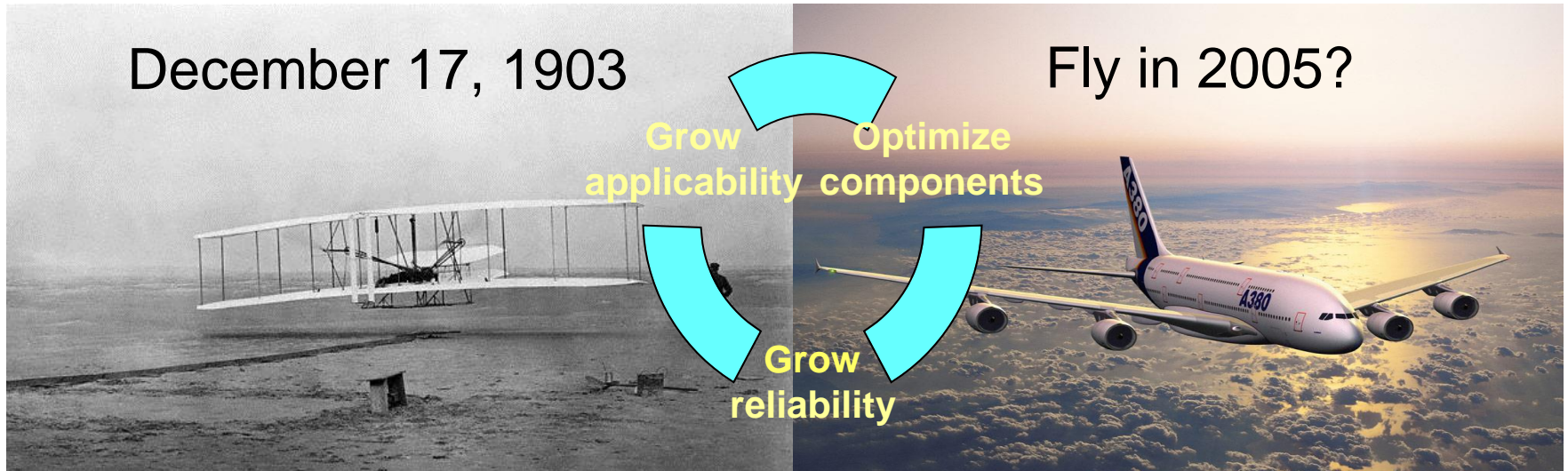
Pave the way towards achieving new functionality
by combining available components



17th December, 1903
12 seconds, 36 meters



Let's make it fly!



100 years of optimization and improvements



Summary

- Standards are evolving
 - Robust enough to build valuable applications today
 - Manage expectations and pay attention to sociology
- Open source software
 - There are GRID Solution providers that can help
- Commercial software licensing models need to evolve
- Skills
 - Those with GRID and Semantic Web skills will be in demand
 - Seek training opportunities (www.nesc.ac.uk)

