

Centers for Disease Control and Prevention

IT Research & Development

Analysis and Design of Organizational Computational Science

Chris Lynberg

April 19, 2010



SAFER • HEALTHIER • PEOPLE™



Outline

- **Analysis**
 - **History**
 - **Ultimate Goals, Value, Metrics, and Challenges**
 - **Requirements**
 - **Organizational Structure**
 - **Computational Biology at CDC**
 - **Types of Computational Science**
 - **Observations**
 - **R&D Strategy and Goals**
- **Design**
 - **Architectural Goals**
 - **Sustainability, Coherency**
 - **Intra and Inter-organizational Continuity**
 - **Technical Goals**
 - **Secure, Flexible, Expandable**
- **Case Studies**
- **Collaborations**

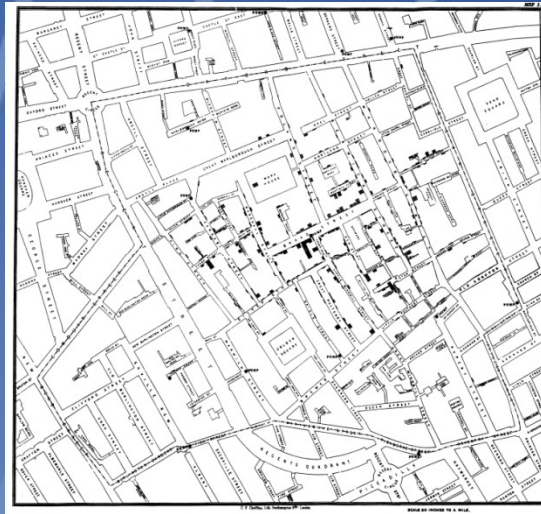


SAFER • HEALTHIER • PEOPLE™



History

John Snow: 1854



Foundational event for Computational Epidemiology.

Requires a synthetic understanding of the underlying model.

Ability to abstractly recognize a pattern without bias.

Predicting the daily weather is easier than predicting disease.

Public Health Science is attributed to lengthening the average lifespan by 25 years since the year 1900.



SAFER • HEALTHIER • PEOPLE™





1942: Office of Malaria Control during WWII

1947: CDC employees purchased campus from Emory for \$10 with Robert Woodruff gift.

1957: Inclusion of STD prevention.

1960: Inclusion of TB prevention.

1963: Immunization Program established.

1980: Centers for Disease Control

1992: Renamed to:
Centers for Disease Control and Prevention

2010: 8,500 FTE's; 15,000 total workforce ; \$6.8B; 50 states; 45 countries.

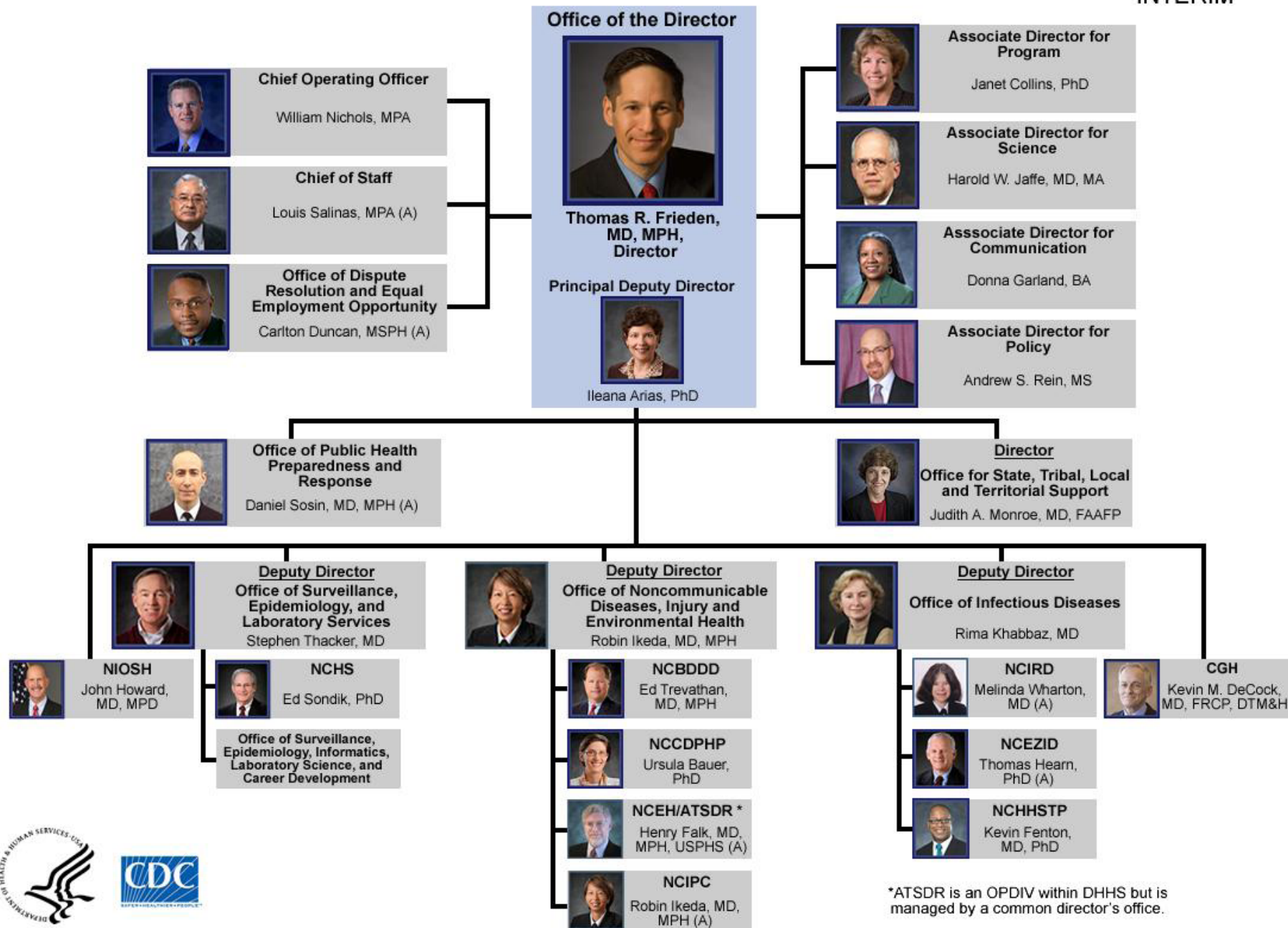


SAFER • HEALTHIER • PEOPLE™



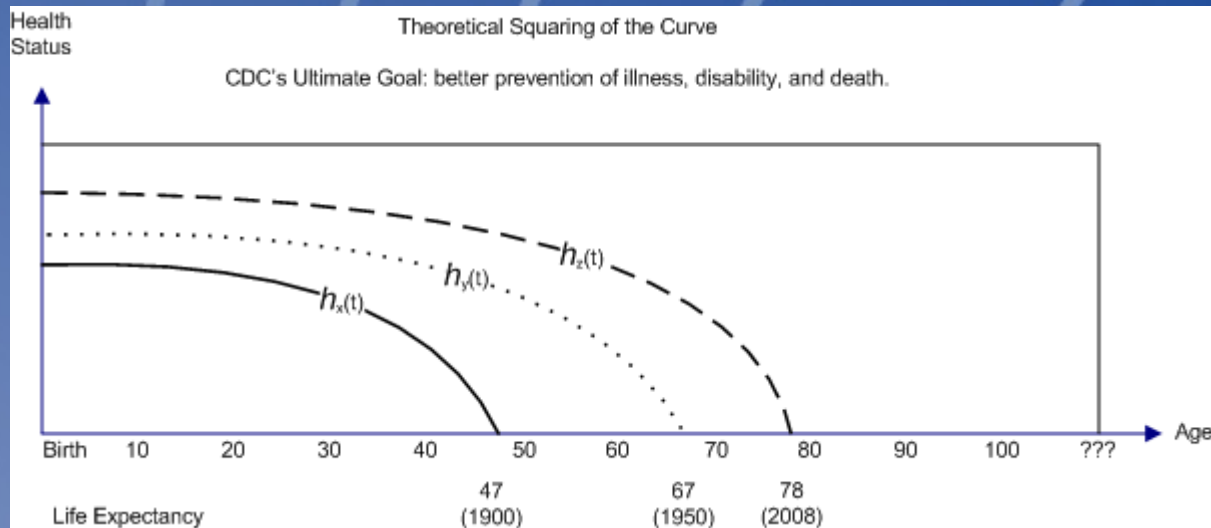
Centers for Disease Control and Prevention

INTERIM



*ATSDR is an OPDIV within DHHS but is managed by a common director's office.





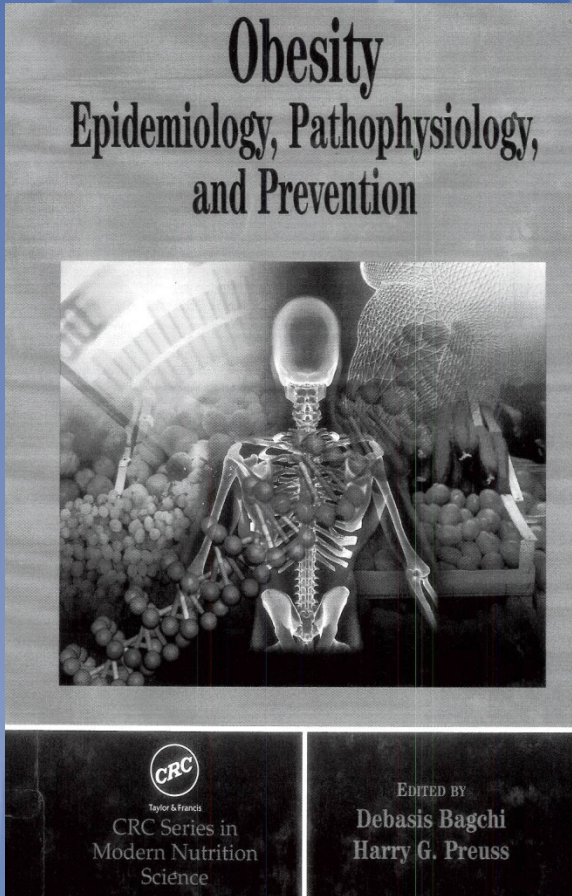
Theoretical national productivity gains if the equations could be defined: $P_g = \int h_z(t)dt - \int h_y(t)dt$



SAFER • HEALTHIER • PEOPLE™



Calorie Restriction in Mice



266

Obesity: Epidemiology, Pathophysiology, and Prevention

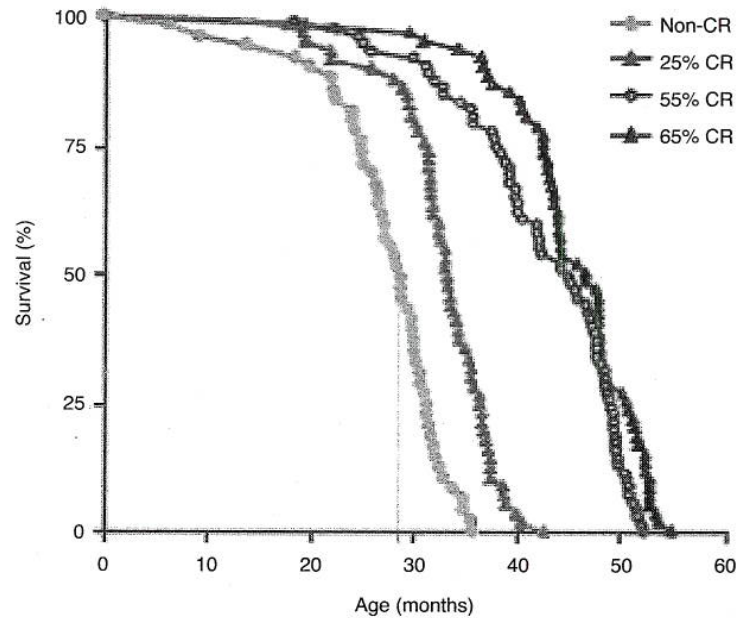


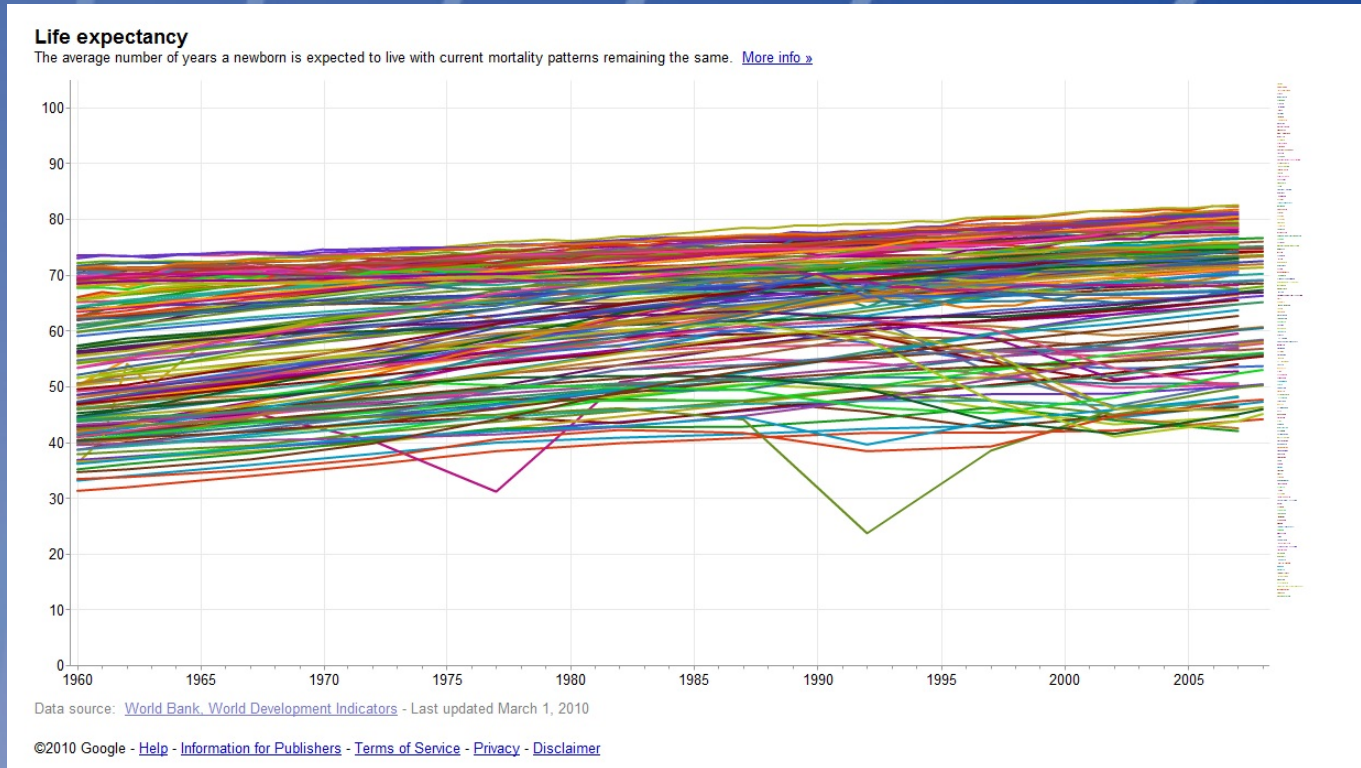
FIGURE 21.1 Effects of increasing levels of caloric restriction on survival in mice. Caloric restriction causes a parallel rightward shift in the survival curve, increasing both average and maximal life span. (From Weindruch, R. et al., *J. Nutr.*, 116, 641–654, 1986. With permission.)



SAFER • HEALTHIER • PEOPLE™



Composite image of Life Expectancy



1960
Sweden(73)
Afghanistan(31)
WorldAvg(50.2)

Gap = 42

2007
Japan(83)
Afghanistan(44)
WorldAvg(68.9)

Gap = 39



SAFER • HEALTHIER • PEOPLE™



Leading Causes of Death in the U.S.A (2006 Data)

- Heart disease: 631,636
- Cancer: 559,888
- Stroke (cerebrovascular diseases): 137,119
- Chronic lower respiratory diseases: 124,583
- Accidents (unintentional injuries): 121,599
- Diabetes: 72,449
- Alzheimer's disease: 72,432
- Influenza and Pneumonia: 56,326
- Nephritis, nephrotic syndrome, and nephrosis: 45,344
- Septicemia: 34,234

Source: Deaths: NCHS Final Data for 2006, table B



SAFER • HEALTHIER • PEOPLE™



Priority Areas of Strategic Focus

United States

- Tobacco Control
- Improving nutrition, physical activity, and food safety
- Reducing healthcare-associated infections
- Preventing motor vehicle injuries
- Teen pregnancy prevention
- HIV prevention

Advancing Global Health

- Reducing mother-to-child HIV transmission
- Cutting malaria illness and death in half in high-burden countries
- Saving the lives of children under age five
- Eliminating Lymphatic Filariasis globally, and Onchocerciasis in the Americas
- Reducing salt intake in China to prevent and control high blood pressure
- Strengthening health ministries to better prevent communicable and non-communicable diseases and injuries.

Malaria is a mosquito-borne disease caused by a parasite. People with malaria often experience fever, chills, and flu-like illness. Left untreated, they may develop severe complications and die. Each year 350-500 million cases of malaria occur worldwide, and over one million people die, most of them young children in sub-Saharan Africa.



SAFER • HEALTHIER • PEOPLE™



Example National Productivity Loss Estimations

- Diabetes: \$58 billion in reduced national productivity in 2007. Economic Costs of Diabetes in the U.S. in 2007. ADA Statement. Diabetes Care 31:596–615, 2008
- Annual Flu: Projected lost earnings due to illness and loss of life amounted to \$16.3 billion annually. The annual impact of seasonal influenza in the US: Measuring disease burden and costs. Vaccine 25 (2007) 5086–5096. Noelle-Angelique M. Molinari, et al.
- Smoking: \$96.8 billion (\$64.2 billion for males and \$32.6 billion for females) Smoking-Attributable Mortality, Years of Potential Life Lost, and Productivity Losses --- United States, 2000—2004. Morbidity & Mortality Weekly Report. 2008;57(45):1226-1228. © 2008 Centers for Disease Control and Prevention (CDC)
- Injury: The cost of injury in the United States also is substantial, exceeding \$400 billion in 2000\$68 billion in productivity losses. Guidelines for Field Triage of Injured Patients Recommendations of the National Expert Panel on Field Triage Morbidity & Mortality Weekly Report. January 23, 2009 / 58(RR01);1-35. © 2008 Centers for Disease Control and Prevention (CDC)
- Cervical Cancer: \$1.3 billion: Estimated productivity loss in 2000 owing to cervical cancer mortality. Womens Health Issues. 2006 Sep-Oct;16(5):236-42. Annual productivity costs due to cervical cancer mortality in the United States. Insinga RP. Department of Health Economic Statistics, Merck Research Laboratories, Merck & Co. Inc., North Wales, PA 19454, USA.
- Challenge: We need valuable, strategic, and mission oriented metrics for all forms of computational science.



SAFER • HEALTHIER • PEOPLE™



Example National Productivity Loss Estimations

United States

Over 162 million cases of seven common chronic diseases — cancers, diabetes, heart disease, hypertension, stroke, mental disorders, and pulmonary conditions — were reported in The United States in 2003. These conditions shorten lives, reduce quality of life, and create considerable burden for caregivers.

Year	Total	Treatment Costs	Productivity Loss
2003	\$1,324B	\$277B	\$1,047B
2023 (in 2003 \$'s)	\$4,130B	\$790B	\$3,363B

Milken Institute analysis based on MEPS, NHIS, etc.

For more information refer to full report “An Unhealthy America: Economic Burden of Chronic Disease” or visit www.milkeninstitute.org



SAFER • HEALTHIER • PEOPLE™



Obesity

Prevention Makes Common Cents, Department of Health and Human Services, 2003.

<http://aspe.hhs.gov/health/prevention/prevention.pdf>

“20-year-old white male could realize a 17 percent reduction in life expectancy due to obesity.
(37)”

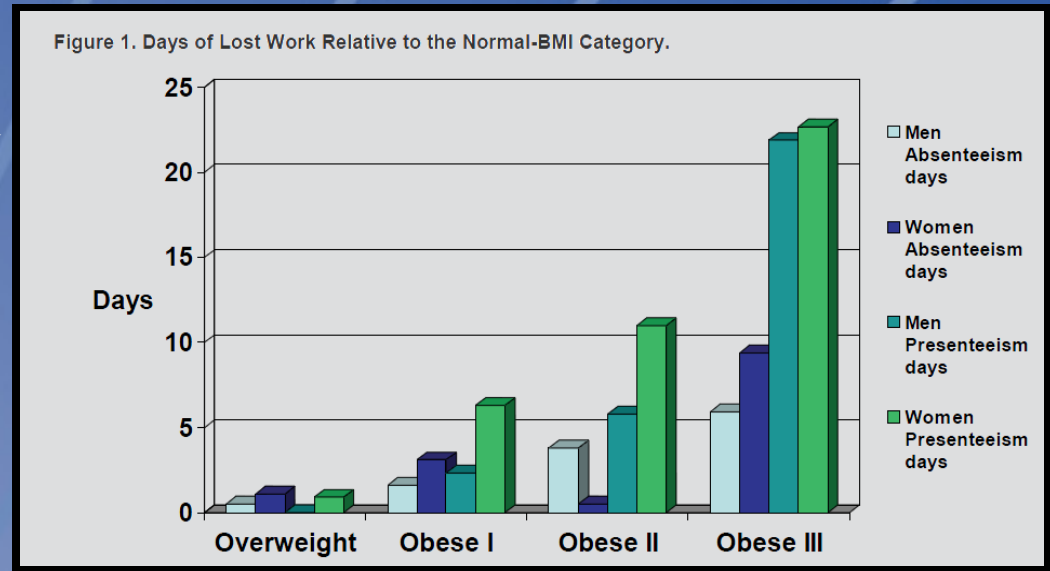
Overweight and obesity raise the risk for:

- type 2 diabetes
- high blood pressure
- high cholesterol levels
- coronary heart disease
- congestive heart failure
- angina pectoris
- stroke
- asthma
- osteoarthritis
- musculoskeletal disorders
- gallbladder disease
- sleep apnea and respiratory problems
- gout
- bladder control problems
- poor female reproductive health
 - complications of pregnancy
 - menstrual irregularities
 - infertility
 - irregular ovulation
- cancers of the
 - uterus
 - breast
 - prostate
 - kidney
 - liver
 - pancreas
 - esophagus
 - colon and rectum

THE ASSOCIATION BETWEEN OBESITY AND PRODUCTIVITY LOSS

Kannan , Finkelstein, Burgess, DiBonaventura , Hale

Consumer Health Sciences, Duke-NUS Graduate Medical School, RTI International, Global Health Outcomes Strategy and Research, Allergan Inc,



<http://www.chsinternational.com/Resources>

[%5C2009_10_Obesity_Conference_Obesity_Productivity_Loss_Allergan_c2854%20-%20Allergan.pdf](http://www.chsinternational.com/Resources/%5C2009_10_Obesity_Conference_Obesity_Productivity_Loss_Allergan_c2854%20-%20Allergan.pdf)

http://en.wikipedia.org/wiki/Obesity_in_the_United_States



SAFER • HEALTHIER • PEOPLE™



An example algorithm to estimate the loss of national productivity.

$$\Delta ADays_{a,g} = L_{a,g} \times \sum_{w,h} \Delta P_{a,g,w,h} \times \left(\overline{ADays}_{a,g,w,h} - \overline{ADays}_{a,g,w-normal,h-no} \right)$$

$$\Delta PDays_{a,g} = L_{a,g} \times \sum_{w,h} \Delta P_{a,g,w,h} \times \left(\overline{PDays}_{a,g,w,h} - \overline{PDays}_{a,g,w-normal,h-no} \right)$$

$$\Delta Disability_{a,g} = \sum_{w,h} \Delta P_{a,g,w,h} \times \left(DisabilityRate_{a,g,w,h} - DisabilityRate_{a,g,w-normal,h-no} \right)$$

$$NPV_{a,g} = \sum_{A-a}^{64} \frac{\left(E_{A,g} \times (L_{A,g} + 0.75 \times [1 - L_{A,g}]) \times (1+p)^{(A-a)} \times [1 - CM_{A,g}] \right)}{(1+d)^{(A-a)}} + \sum_{A-65}^{74} \frac{\left(E_{A,g} \times L_{A,g} \times (1+p)^{(A-a)} \times [1 - CM_{A,g}] \right)}{(1+d)^{(A-a)}}$$

$$\Delta Mortality_{a,g} = \sum_{w,h} \Delta P_{a,g,w,h} \times \left(MortalityRate_{a,g,w,h} - MortalityRate_{a,g,w-normal,h-no} \right)$$

$$\Delta National\ Productivity = \sum_{a,g} \left(\begin{array}{l} \Delta ADays_{a,g} \times \overline{DailyEarnings}_{a,g} \\ + \Delta PDays_{a,g} \times \overline{DailyEarnings}_{a,g} \\ + \Delta Disability_{a,g} \times \overline{AnnualEarnings}_{a,g} \\ + \Delta Mortality_{a,g} \times NPV_{a,g} \end{array} \right)$$

“If there were no cases of excess weight and hypertension, we estimate that national productivity would be \$196 billion higher than current levels, indicating the maximum potential benefits achievable by eliminating these risk factors (Table 2). Specifically, overweight and obesity are associated with a \$144 billion decline in national productivity, of which 63% is attributed to obesity.”

Predicted National Productivity Implications of Calorie and Sodium Reductions in the American Diet.

Timothy M. Dall, MS; Victor L. Fulgoni III, PhD; Yiduo Zhang, PhD; Kristin J. Reimers, PhD;

Patricia T. Packard, MS, RD; James D. Astwood, PhD. American Journal of Health Promotion. July/August 2009, Vol. 23, No. 6



SAFER • HEALTHIER • PEOPLE™



Life Saving Science Articles at CDC

January 25, 2010

First Meeting of the Genomic Applications in Practice and Prevention Network

January 11, 2010

Diabetes mellitus and birth defects

October 5, 2009

CDC's Technology Transfer Office (TTO) works to translate CDC research into commercial products that improve public health.

June 29, 2009

July is National Hemochromatosis Awareness Month

June 8, 2009

CDC's Screen for Life: National Colorectal Cancer Action Campaign

May 25, 2009

Pool Chemical Injuries Send Thousands to Emergency Department Each Year

May 11, 2009

May is Asthma Awareness Month

May 4, 2009

Improving the Nation's Vision Health

April 6, 2009

Reducing Beryllium Sensitization Among Workers

March 31, 2009

Promoting Oral Health in Children

March 24, 2009

Fetal Alcohol Syndrome Disorders: Prevention, Identification, & Intervention

March 17, 2009

Combating XDR TB

March 10, 2009

Advancements in measurements of toxins using MS



SAFER • HEALTHIER • PEOPLE™



Life Saving Science Articles at CDC

- March 3, 2009
Preventing, Responding & Recovering from Brain Injury
- February 27, 2009
Science on a mission: Science-related workgroups
- February 20, 2009
Cool tool to keep-up with a child's immunizations
- January 23, 2009
Communicating about Chagas disease in the US
- January 16, 2009
Tracking diabetes one county at a time
- January 9, 2009
CDC works to understand, prevent birth defects
- January 2, 2009
Novel science through Planning Grants
- December 8, 2008
Simulating Health System Transformation by Playing CDC's Health Protection Game
- December 1, 2008
Focus on perinatal HIV prevention in the US
- November 24, 2008
Finding new ways to protect against polio
- November 17, 2008
In CDC's public health activities, info is king
- November 11, 2008
Nanotechnology and Public Health



SAFER • HEALTHIER • PEOPLE™



The National Importance of HPC

Federal reports and methods of advancing science:

- High-End Computing Revitalization Task Force, 2004
- NAS Report on the Future of Supercomputing, 2005
- PITAC Report on Computational Science, 2005
- Cyberinfrastructure Vision for 21st Century Discovery, 2007
- Agency Performance: Federal Enterprise Architecture



SAFER • HEALTHIER • PEOPLE™



Areas of Computational Science

Prevention Modeling

Influenza Pandemic Preparation

Molecular Modeling

Distribution Modeling

Data Analysis

Monte Carlo Analysis

Computational Toxicology

3-D modeling and animation

Computational Fluid Dynamics

Virtual Reality Development

Nanotechnology

Population Genetics

Genome annotation

Computational Genomics

Genetic Analysis

Comparative Genomics

Phylogenetic Analysis

Phylogenetic Construction

Sequence Analysis

Protein Structure Prediction

Surveillance: pre, current, post
Syndromic Surveillance.



SAFER • HEALTHIER • PEOPLE™



Computational Biology

Computational Epidemiology

Computational Biomodeling

Bioinformatics

Mathematical Biology

Computational Genomics

Molecular Modeling

Protein Structure Prediction

Computational Biochemistry and Biophysics



SAFER • HEALTHIER • PEOPLE™



Computational Biology Focus



The IEEE/ACM Transactions on Computational Biology and Bioinformatics emphasize the algorithmic, mathematical, statistical, and computational methods that are central in bioinformatics and computational biology ... where many forms of data and biological knowledge are used to create a computer-based model of a complex biological system.



The International Society for Computational Biology (ISCB) is an organization dedicated to the advancement of scientific understanding of living systems through computation. 2000 members; 50 countries.

Challenge: Finding computational biology technologies that will help meet CDC's mission and goals.



SAFER • HEALTHIER • PEOPLE™



Requirement for Language Clarity

- HPC (High Performance Computing): Systems comprising of very fast resources, typically 100's or 1000's of processors, and very fast memory, network, and storage.
- Scientific Computing: Applications addressing a particular field of science, not necessarily requiring HPC resources.
- Computational Science: Science done by computations rather than by theory and experiment alone, which typically requires HPC resources.



SAFER • HEALTHIER • PEOPLE™



Types of Computational Science

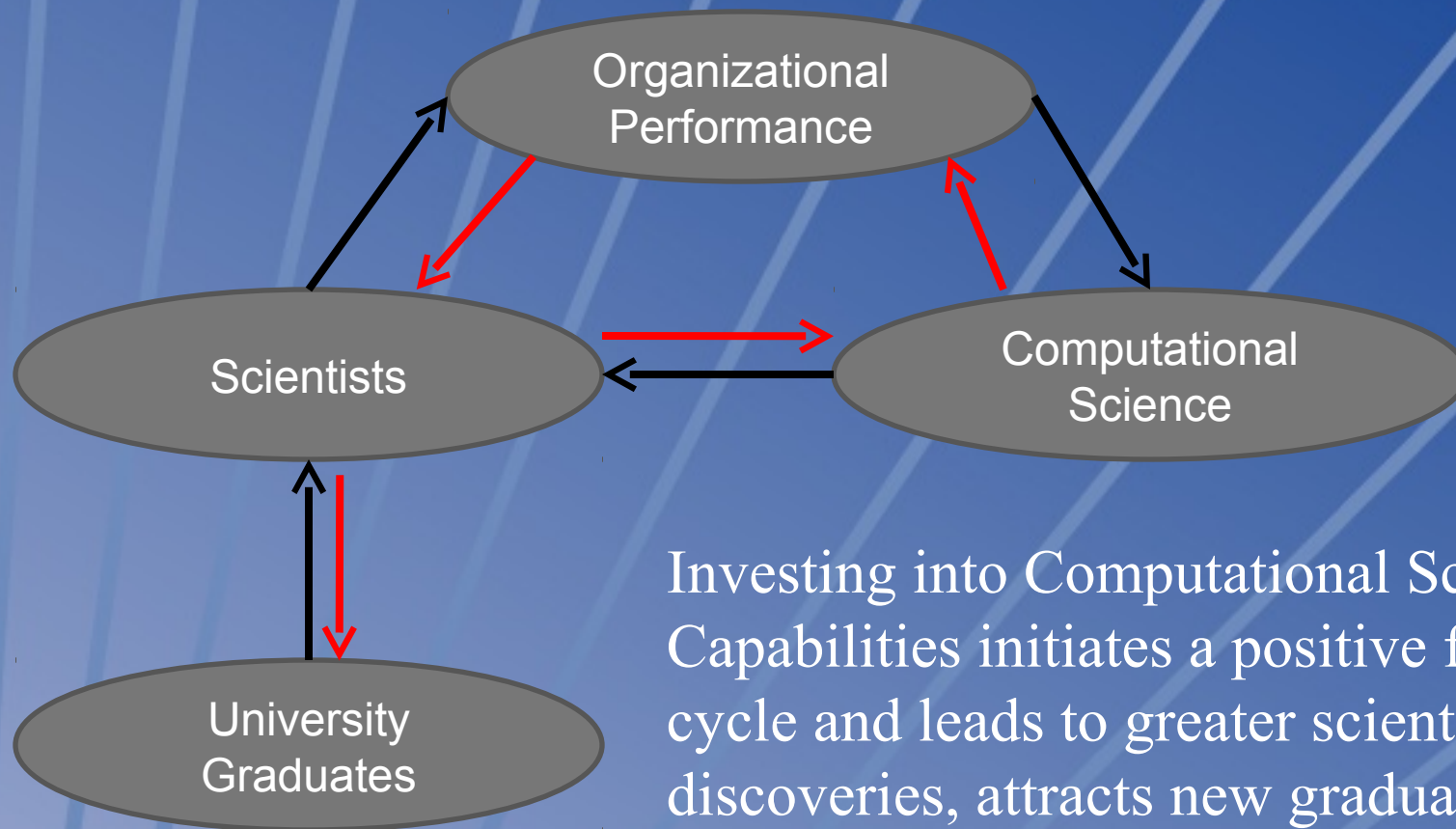
- Repeatable production computing:
 - Daily Weather Predictions
 - Bioinformatics Sequencing of Pathogens
- R&D or “test tube” computational science:
 - Do changes in the Bearing Strait affect Global Warming?
 - Which genetic substitutions have the greatest impact on Hepatitis C evolution?
- These represent the two major domains of computational science at CDC.



SAFER • HEALTHIER • PEOPLE™



Feedback Cycle Observations



Investing into Computational Science Capabilities initiates a positive feedback cycle and leads to greater scientific discoveries, attracts new graduates, and improves agency performance.



Strategic R&D Direction

- Use the Federal Enterprise Architectural Strategy: Architect, Invest, Implement. Blueprint was the first step.
- Win-Win: Speed the time-to-science AND Optimize IT Resources.
- Advance CDC's Computational Science Capabilities
- Federation, collaboration of scientific data that can be used for computational science.



SAFER • HEALTHIER • PEOPLE™



Architectural Formalization of Computational Science

- Changes the organizational perspective.
- Provides a focus on computational science.
- Provides a foundation upon which computational science decisions can be made.



SAFER • HEALTHIER • PEOPLE™



Additional Metrics Required

- Registry of computational science activities:
 - Name of Investigation, Path to Program and Data, Total Compute Hours
 - Publication, Total Funding Required, Resulting Policy, Projected ROI
- Computational Science Metric Standards
 - Publish which systems were utilized.
 - Publish the resources required: Hours, processors, data, archival.



SAFER • HEALTHIER • PEOPLE™



CDC Collaboration

- With a computational science architecture the following types of collaborations are catalyzed:
 - Scientific within CDC
 - Academic external to CDC
 - Coherent Data Architecture and reuse of data.
 - Businesses which rely upon the dissemination of CDC data and information.
 - Scientific Computing Industry
 - Synergy vs. Independence
 - Visible synergy is observable with current case studies.



SAFER • HEALTHIER • PEOPLE™



Initial R&D Activities

- Interviewed scientists with significant scientific computing problems.
- Initiated Internal and External Collaborations
- CDC R&D HPC Lab Research for Emerging Technologies.
- CDC HPC Symposium
- Survey of Computational Science Requirements
- Biotechnology Core Seminars with Emory, GA Tech
- Biotechnology Core seminar on MatLab-GPU-Accelereyes
- Began the process and strategy to develop the CDC framework to write a blueprint towards modernization.



SAFER • HEALTHIER • PEOPLE™



Goals of the Research and Development

- **Research the requirements, issues and parameters:**
 - Outlined in the strategic framework, and proposal.
- **Research Test Cases**
 - Practical understanding of the issues.
 - Exploration of test resources.
- **Research of Emerging Technologies:**
 - Hardware, Software, and Applications.
- **Research Successful Role Models of HPC.**
 - Primarily Fed. Labs with 10-20 yrs experience.

Develop an Architectural Blue Print.



SAFER • HEALTHIER • PEOPLE™



HPC Architecture

Model / Project	P1	P2	P3	P4	P5	P6	P7	P8	P9
Independent	X								
Center/Divisional		X							
CDC			O						
NIH				X					
Other Gov Collaboration					X				
University Collaboration						X			
Terragrid							X		
NCSA								X	
Outsourced									X

June 2008



SAFER • HEALTHIER • PEOPLE™



Elements of the Blueprint

Frame work of sub-architectures:

- Design Goals, and Requirements
- Scientific Strategy, Metrics, and Value
- Business Architecture, Performance, Metrics
- Governance Model
- Security Architecture
- Data Architecture
- Educational Architecture
- Computational Science: models, algorithms, state-of-art
- Resource Driven Architecture
- Scientific Collaborative Leadership
- Legal Architecture: FOIA, HIPPA, eDiscovery, Records
- Information Technology: Systems, Network, Storage, etc...



SAFER • HEALTHIER • PEOPLE™



Requirement for Coherency

(Symbiotic Organelles of Computational Biology)

- Science: (Theory; Experiment)
 - Metrics, Business, Value, Strategy
 - Data Collection, Analysis, Dissemination
- Computational Science: (Algorithms)
 - Performing science computationally.
 - Matching the algorithm to the computer architecture.
- Computer Science: (Silicon, O/S, Programming)
 - How to accelerate Computational Science
 - How to reduce the barriers of parallelization
- Information Technology: (Service Oriented)
 - Cyberinfrastructure, Security, SLA, Management, Help Desk
 - Data Architecture
- Business: (Sustainability)
 - Organizational, Governance, Financing, Policy, Legal, Educational
 - Marketing



SAFER • HEALTHIER • PEOPLE™



Cyberinfrastructure

- Cyberinfrastructure is an IT solution to enable the derivation of novel scientific theories and knowledge.
- Computational Science is used throughout CDC.
- The CDC Resource Index indicates there about 1942 scientific activities and subjects addressed at CDC that will all probably employ computational science in the future.
- Geometrically expanding demands for computations and data storage.
- Every computational science algorithm has custom requirements.
- How to engineer a coherent organizational data architecture?
 - Collaboration and Synergy vs. Legal, Data Security, and Records Management.



SAFER • HEALTHIER • PEOPLE™



Architectural Goals

- Sustainability
 - Metrics, Value, Business model.
- Coherency
 - Harmonizing: Science, Computational Science, Computer Science, IT, Business.
- Continuity
 - Seamless scalability from workstation, server, cluster, enterprise to national lab.
- Improve institutional data architecture
 - Technological foundation, standards, metrics, records, transparency.
- Promote collaboration and synergy
 - Collaborative Leadership, clarity of vision and value.
- CDC goals, and to prepare for future national events.
 - “Better prevention of illness, disability, and death”
- Improve efficiencies, scale of economies, synergy.
 - Reduce administration, enable cross-correlation of data (metadata analysis)



SAFER • HEALTHIER • PEOPLE™



Example Case Studies

- Bioinformatics Systems – Dr. Elizabeth Neuhaus
- NCCD Human Papilloma Virus – Dr. Donatus Ekwueme
- NCID Hepatitis C mutation modeling – Dr. James Lara, Zoya Dimitrova
- NCID MicrobeNet – Dr. Ainsley Nicholson
- NCID H1N1 – Dr. John Glasser
- NCID Hepatitis (Genomic Sequencing database) - Dr. Mike Purdy
- ATSDR GIS – Dr. Stephanie Foster, Andrew Dent
- NCIRD H5N1 Modeling – Dr. Rubin Donis (Dr. Henry Wan)
- ATSDR Groundwater – Morris Maslia, Renee Suarez-Soto
- NCEH Radio nucleotides – Dr. Robert Jones
- NIOSH FLAC3D Mining Modeling - Dr. Mark Larson
- ATSDR Autism – Dr. Eugene Demchuk
- NCCD Spatial Science – Greg Schwartz
- Public Health Genomics – Dr. Muin Khoury
- OSELS Computational Epidemiology – Dr. Tom Savel
- NBDDD Monitoring Birth Defects Surveillance Data – Mike Atkinson



SAFER • HEALTHIER • PEOPLE™



Case Study – Bioinformatics Systems

Authority at CDC is Dr. Elizabeth Neuhaus
Ph.D. in Biochemistry, fluent in IT, Linux, etc...
Collaborative Leader
Arranges for CDC-wide Bioinformatics seminars.
Primary collaborator for the R&D of HPC and Blueprint.
Poster presentation at SC09 with Cornell regarding HCV
Designed the next HPC system for Flu



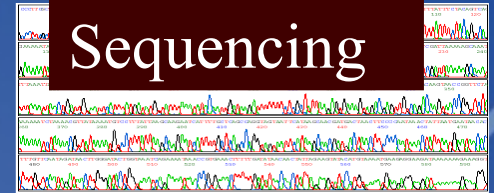
SAFER • HEALTHIER • PEOPLE™



Genome Sequencing for Public Health – Dr. Neuhaus

- Molecular Evolution of Pathogenicity
- Total Microbe Population Analysis
- Total Genome Comparison
- Metagenomics Studies
- Genome Data Mining
- Pathogen Discovery
- Biomarker Discovery

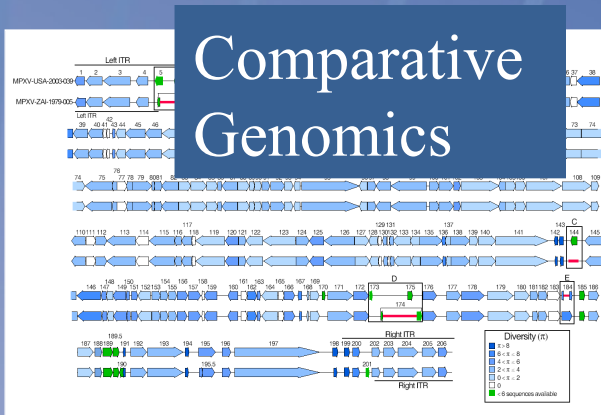
Genome Sequencing



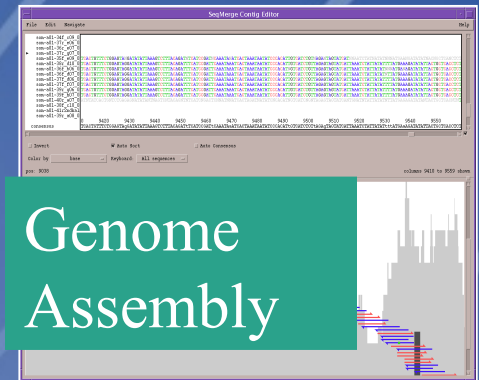
Molecular Evolution



Comparative Genomics



Genome Assembly

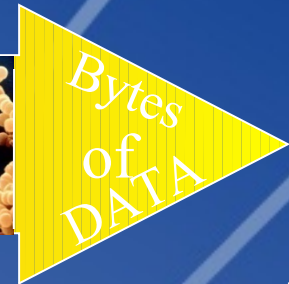
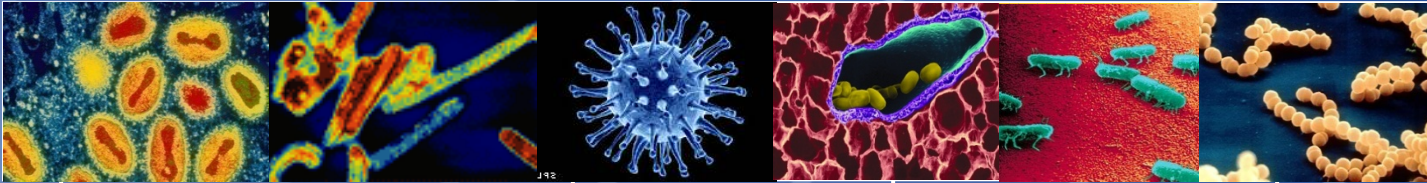


SAFER • HEALTHIER • PEOPLE™



CDC Genome Sequencing – Dr. Neuhaus

High DATA Generation and Special Computing Requirements



2000

2004

2006

2009

Initial Projects

- Human smallpox
- Vaccinia
- CowPox
- MonkeyPox

Total genomes: 95

Hemorrhagic Fevers

- Ebola
- Rift Valley
- Marburg
- Crimean-Congo

Total genomes: 160

Influenza Viruses

- H5N1
- H1N1
- H1N2
- B

Total genomes: >800

454 FLX pyrosequencing

- Anthracis
- Legionella
- Bartonella
- Salmonella
- Neisseria Meningitidis
- Burkholderia
- Coxiella
- Brucella

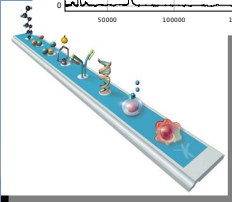
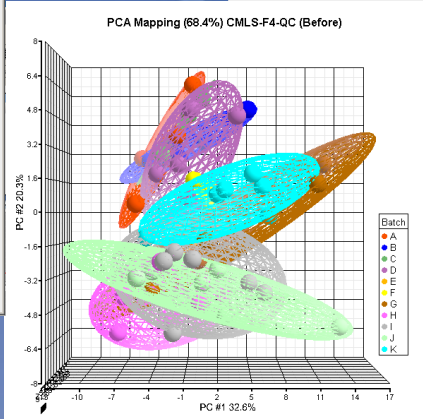
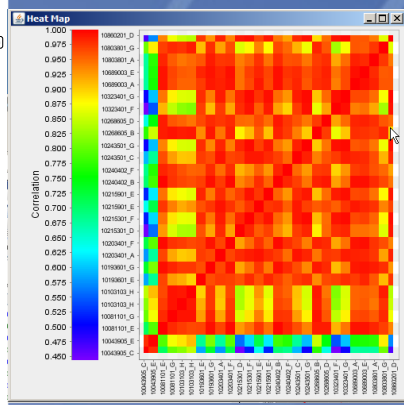
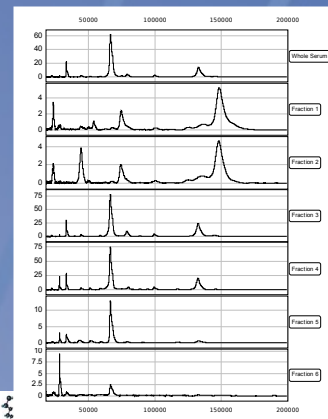
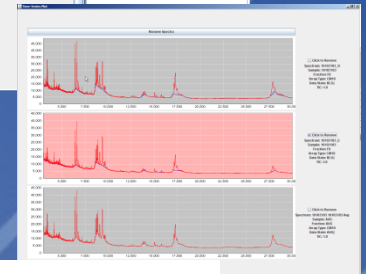
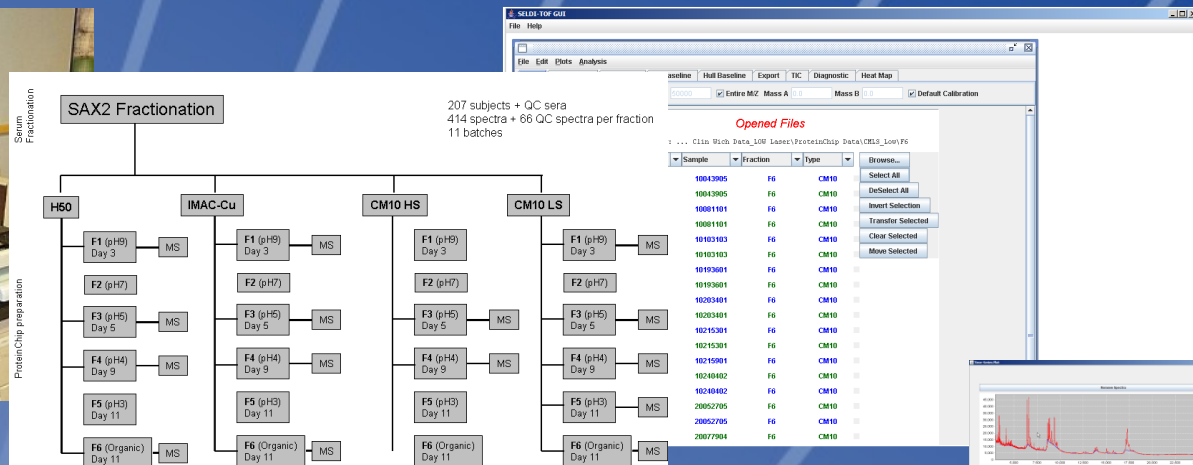
Total bacterial genomes: 45



SAFER • HEALTHIER • PEOPLE™



Proteomics – Dr. Neuhaus



SAFER • HEALTHIER • PEOPLE™



Case Study – Human Papilloma Virus

Probabilistic Simulation Model:

Written in C and MPI.

Estimated to take 675,000 compute hours.

Developed by Drs. Ekwueme, Saraiya, Geng, Franco, and Brisson.

Dr. Ilene Carpenter at ORNL is reviewing the code.

Goal:

Estimate the long-term impact of the Human Papilloma Virus (HPV) vaccination on disease burden and cost for future National Breast and Cervical Cancer Early Detection Program (NBCCEDP) participants. The model will inform the design of optimal screening strategies for the NBCCEDP participants in the presence of HPV vaccination.

Abstract:

Existing studies have not modeled impact of the vaccines on low-income, medically underserved populations in the US and therefore did not account for potentially population heterogeneity. The model belongs to a new generation of individual-based, micro-simulation models that provides the opportunity to estimate the impact of interventions in populations with mixed combinations of characteristics that affect screening and vaccination outcomes. The project will help fill gaps in knowledge of the long-term impact of HPV vaccination in medically underserved and low-income populations in the US.



SAFER • HEALTHIER • PEOPLE™



Case Study – Hepatitis C Virus

Zoya Dimitrova

Developed MatLab algorithm addressing coordinated mutations that evolve with HCV.

Using multiple computational architectures to find optimal solution.

Challenge: Having a library of parallelized algorithms for the right computer architecture.

Dr. James Lara

Support to state and national agencies for the prevention and control of viral hepatitis.

Molecular epidemiology of viral hepatitis such as outbreak investigations and epidemiology of transmission has resulted in a copious amount of data.

454 sequencing and alignment of hundreds of thousands (>400,000) sequence variants using exact or heuristic algorithms requires high performance computing.

3D structure templates are not available for rational design of peptides and proteins to aid in our diagnostics.

Compute bound Bayesian networks for Molecular epidemiological studies.

New computational technologies, services and development/application of faster algorithms will be necessary in the very near future to analyze and process these huge amounts of data.

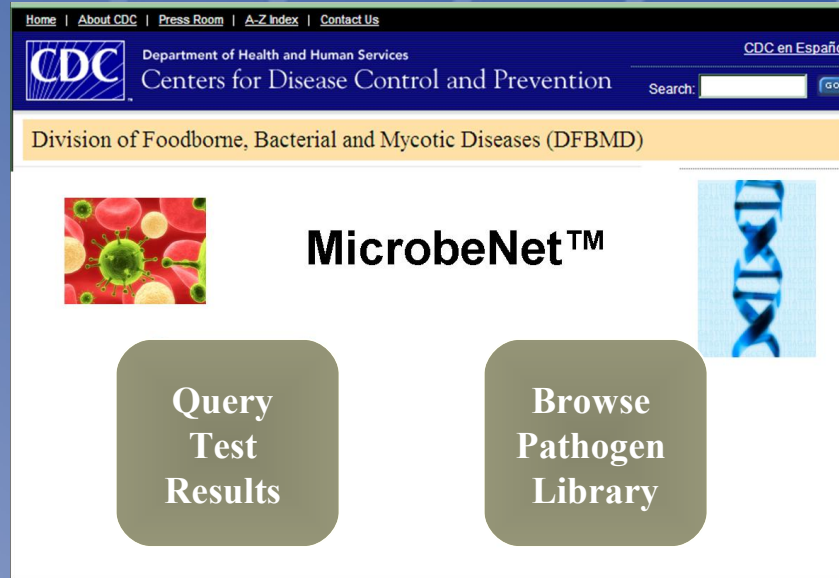


SAFER • HEALTHIER • PEOPLE™



MicrobeNet™

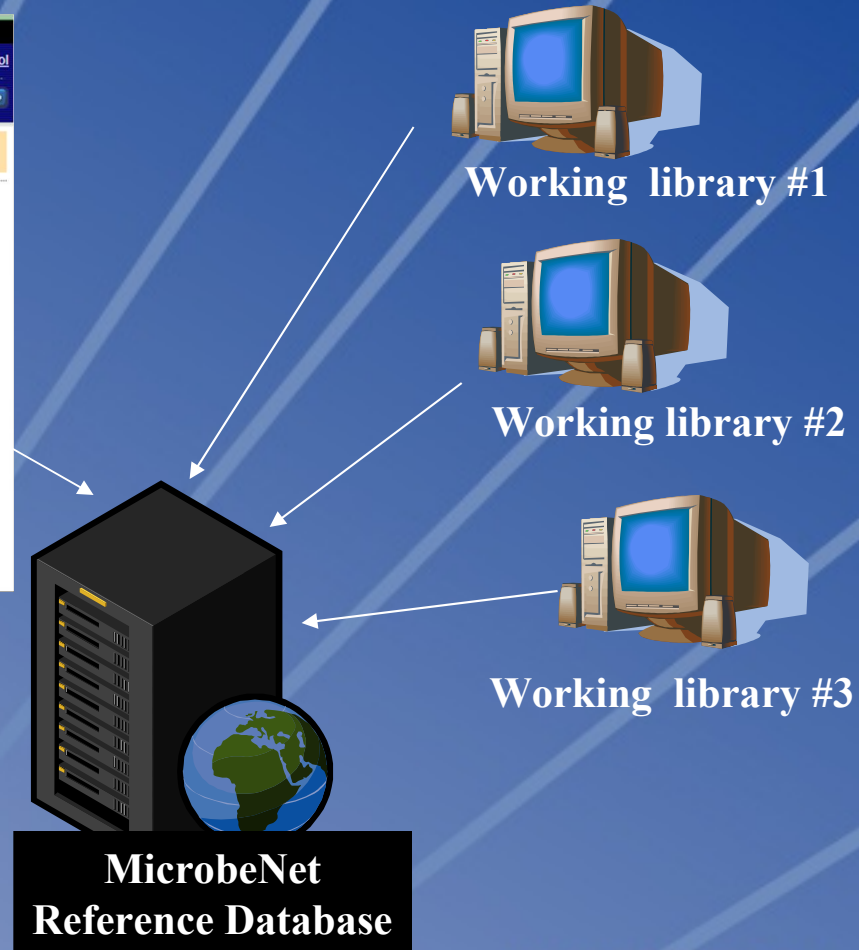
A unified, secure portal for pathogen identification resources.



Question



Answer



SAFER • HEALTHIER • PEOPLE™



Case Study - MicrobeNet

- NCID MicrobeNet - Dr. Ainsley Nicholson
 - Pathogen Identification

The strategic importance of CDC's ability to rapidly and accurately identify pathogens cannot be overstated. This expertise is essential for identifying emerging infectious agents, detecting bioterrorism events, and making informed medical and public health decisions. With our reference laboratories, and collections of tens of thousands of clinical isolates, CDC is recognized world-wide as *the* authority on the phenotypic and genotypic testing of pathogens. Scientists CDC-wide are collaborating to apply computational science technologies to pathogen discovery and identification. To support our state and local public health partners in identification of recognized and emerging pathogens, we are creating MicrobeNet, a highly curated set of tools and data that will be a unique and trustworthy "one-stop shop" source of pathogen identification information. *The next strategic step for MicrobeNet will be to use computational science to improve and accelerate the algorithms that analyze and process genetic and phenotypic data, providing the public with even more accurate, precise, and timely responses.*

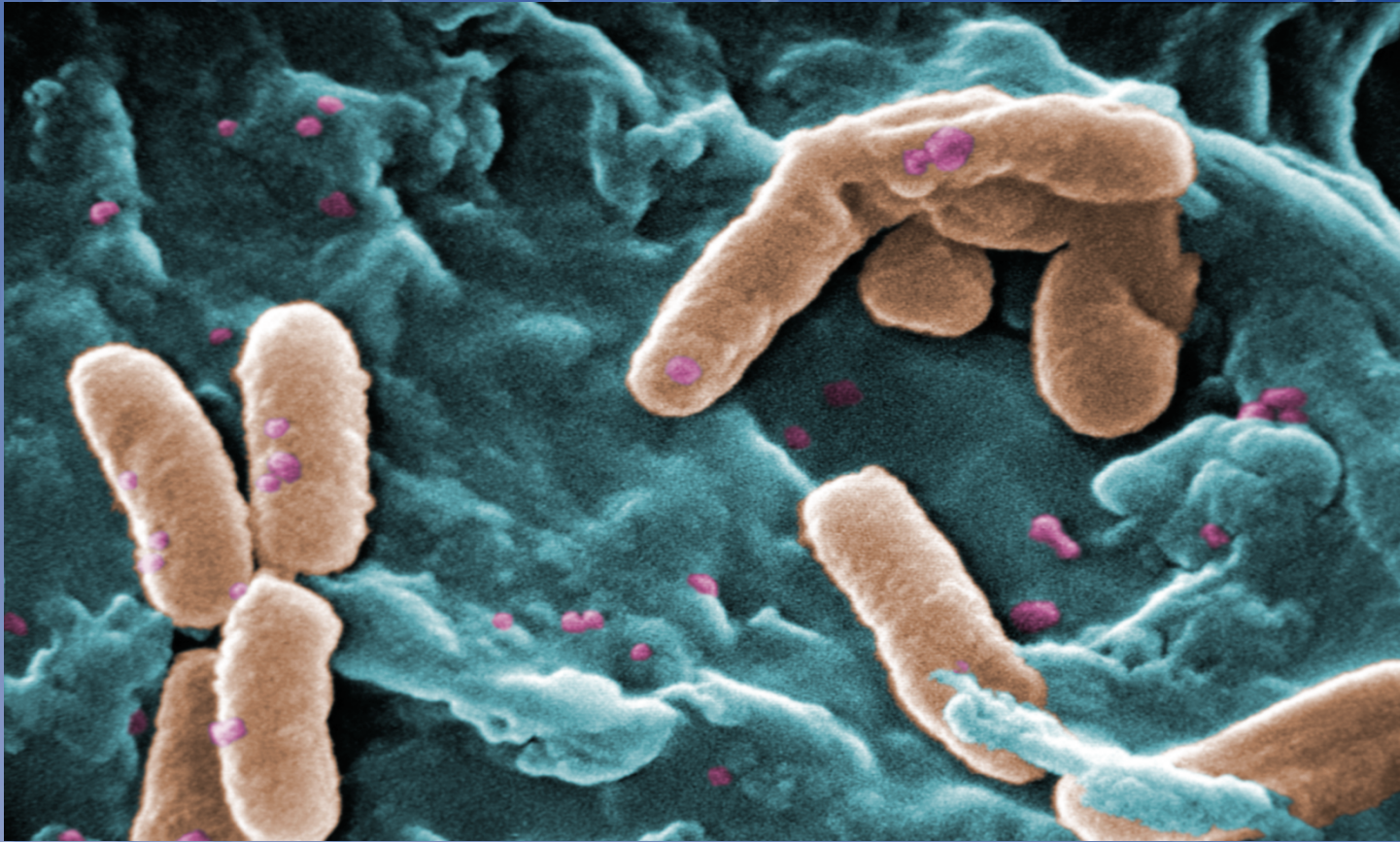


SAFER • HEALTHIER • PEOPLE™



A MicrobeNet Challenge

Example Historic Hand Written Card of *Pseudomonas aeruginosa*.



MicrobeNet-Card-9-24-2009-4-12-40 PM.pdf

SAFER • HEALTHIER • PEOPLE™



Federal Collaborations

- NITRD/HEC - Dr. Robert Bohn, Dr. Chris Greer
- DoE – William Turnbull
 - ORNL – Drs. Julia White, April McMillan
- DoD – Cray Henry
- Army – Dr. Jacques Reifman
- NASA – Bryan Biegel
- NOAA – Mike Kane
- NARA - Ken Thibodeau
- NIH - Steve Bailey (HPC), Jack Collins (GPU)
- NCSA - Dr. Ian Brooks, Dr. Thomas Dunning, Dr. John Towns



SAFER • HEALTHIER • PEOPLE™



University Collaborations

- GA Tech – Dr. David Bader, Dr. Jeffery Vetter
- Cornell University – Dr. David Lifka
- Louisiana State University – Dr. Gabrielle Allen
- Rensselaer Polytechnic Institute – Dr. Fran Berman
- Emory – Dr. Joel Saltz



SAFER • HEALTHIER • PEOPLE™



Corporate R&D

- Intel
- NVIDIA
- Microsoft, Red Hat, Sun
- Cisco
- Accelereyes, Mathworks, Wolfram
- Dell, HP, IBM, JRT
- TerreMark, Amazon



SAFER • HEALTHIER • PEOPLE™



Questions?



SAFER • HEALTHIER • PEOPLE™

