# Thoughts on "The Future of Advanced Cyberinfrastructure for Science and Engineering Research and Education"

Ed Lazowska

Bill & Melinda Gates Chair in
Computer Science & Engineering

Founding Director, eScience Institute

University of Washington

National Science Board

September 2013

http://lazowska.cs.washington.edu/NSB.pdf

# This morning …

- Why must America remain the world leader in computer science?
- How did we gain the lead, and how can we retain it?
- How should our competitiveness be defined?
- The coming decade: Dramatic improvements in technology and algorithms enable "smart everything"
- Cyberinfrastructure to support 21st century "smart discovery"
  - Implications for academia
  - Implications for research policy
  - Implications for K-12 education

# REPORT TO THE PRESIDENT AND CONGRESS

# DESIGNING A DIGITAL FUTURE: FEDERALLY FUNDED RESEARCH AND DEVELOPMENT IN NETWORKING AND INFORMATION TECHNOLOGY


Executive Office of the President

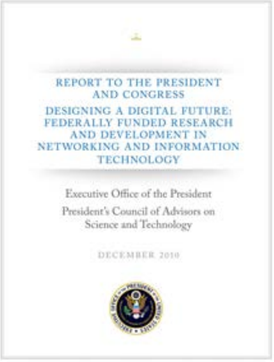
President's Council of Advisors on Science and Technology



DECEMBER 2010

# Why must America remain the world leader in computer science?

- "A key driver of economic competitiveness"
- "Crucial to achieving our major national and global priorities in areas such as energy and transportation, education and life-long learning, healthcare, and national and homeland security"
- "Accelerates the pace of discovery in nearly all other fields"
- "The dominant factor in America's science and technology employment"
- An intellectual agenda "as rich as that of any other field of science or engineering"



REPORT TO THE PRESIDENT AND CONGRESS

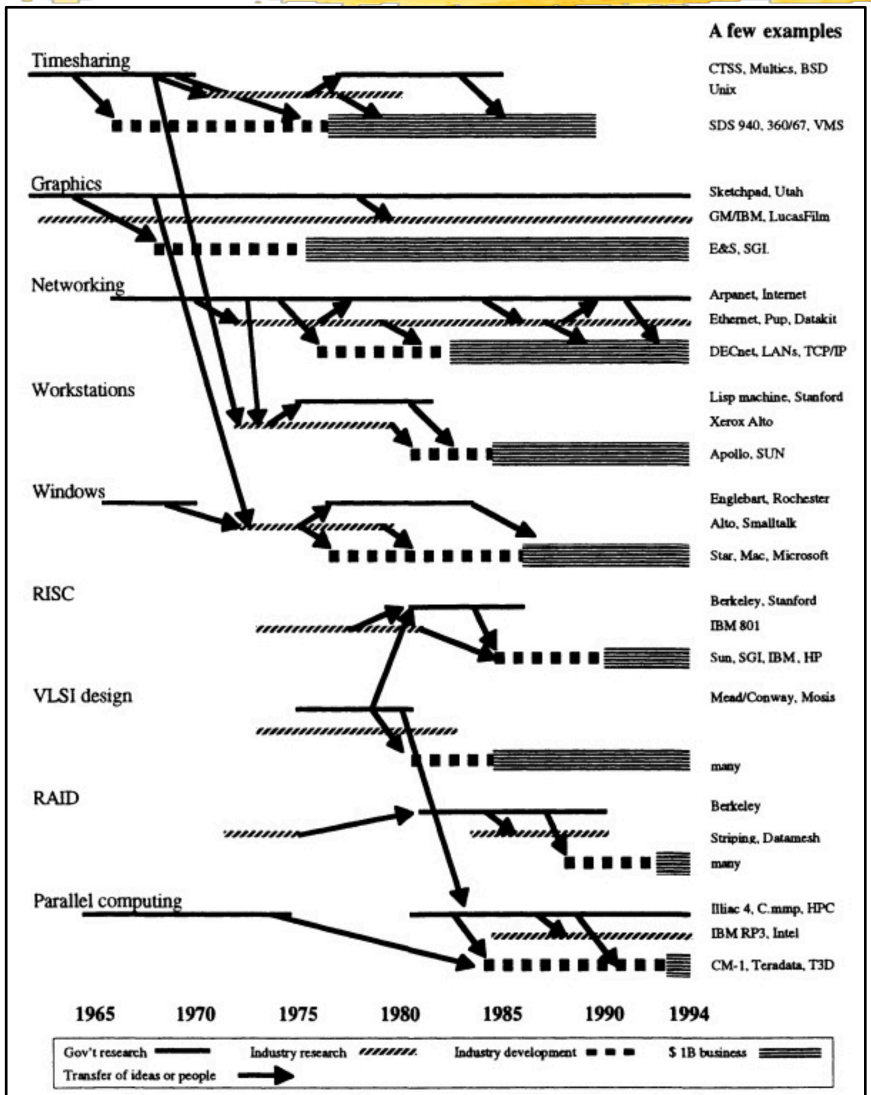
DESIGNING A DIGITAL FUTURE: FEDERALLY FUNDED RESEARCH AND DEVELOPMENT IN NETWORKING AND INFORMATION TECHNOLOGY

Executive Office of the President

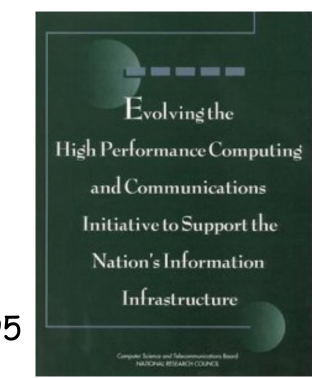
President's Council of Advisors on Science and Technology

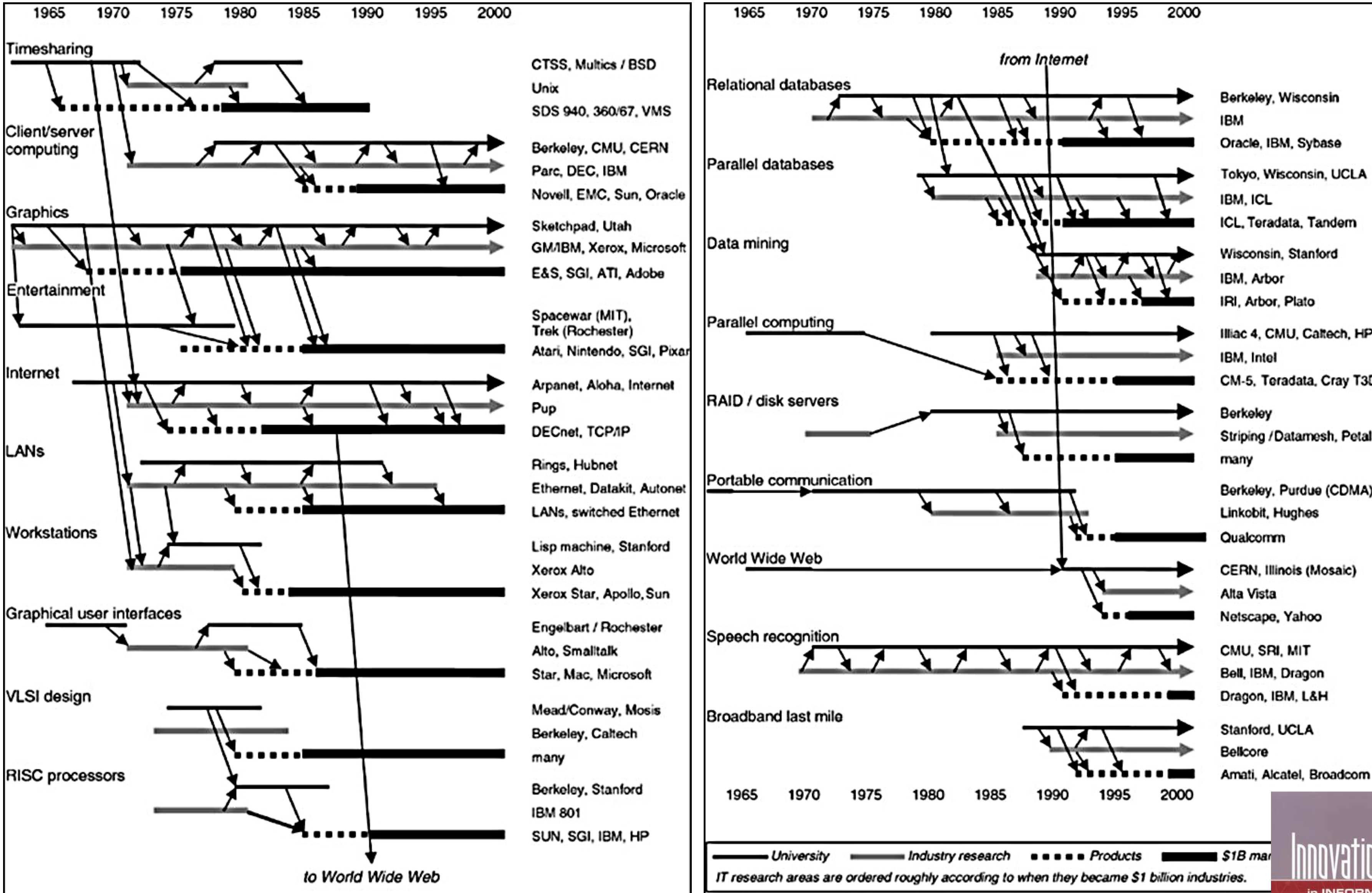
DECEMBER 2010

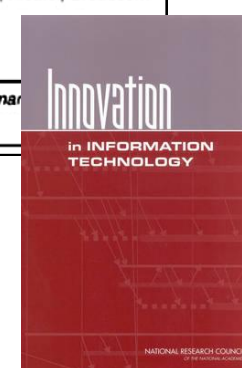
# How did we gain the lead, and how can we retain it?



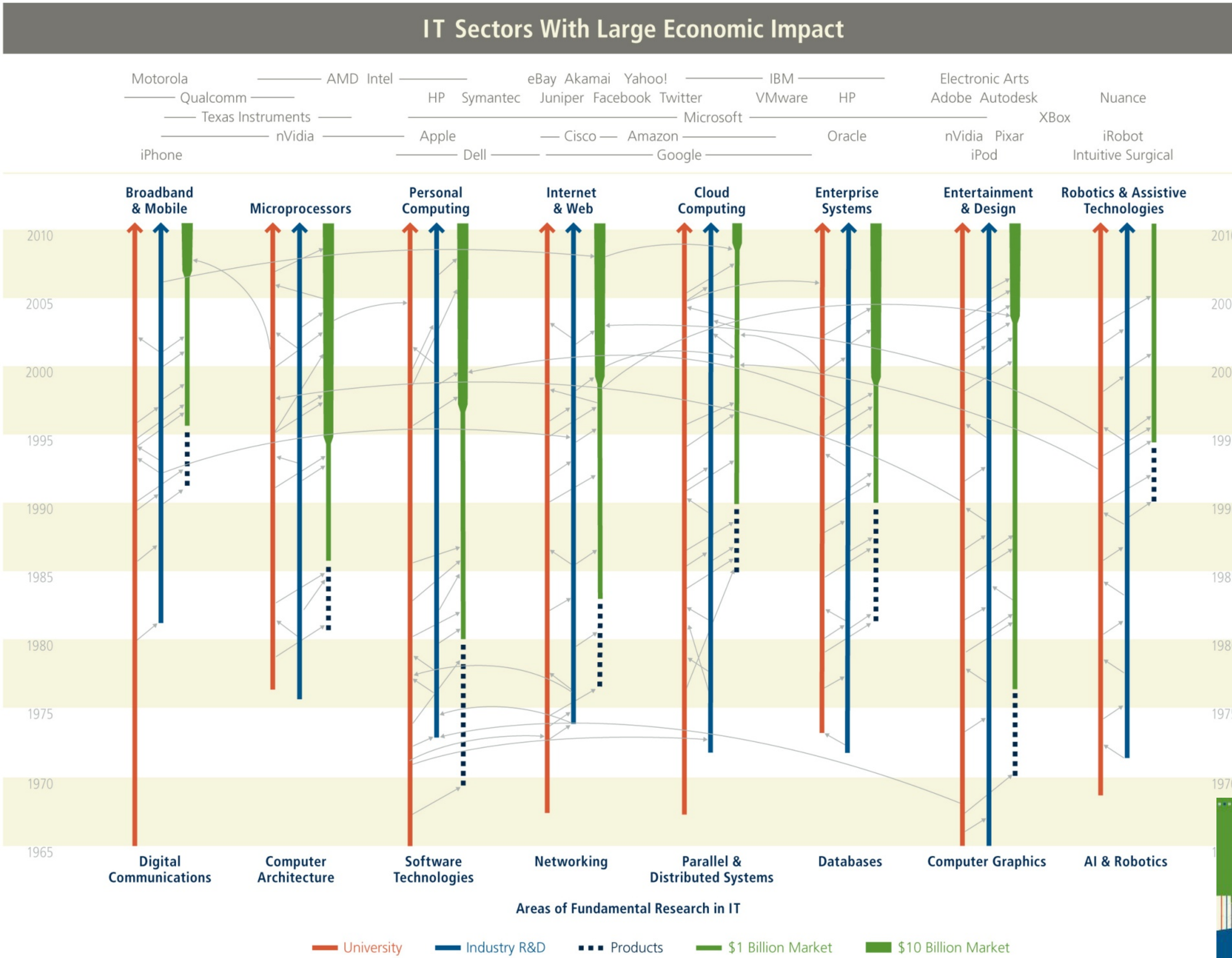

1995




1965
1970
1975
1980
1985
1990
1995
2000
Timesharing
CTSS, Multics / BSD
Unix
SDS 940, 360/67, VMS
Client/server computing
Berkeley, CMU, CERN
Parc, DEC, IBM
Novell, EMC, Sun, Oracle
Graphics
Sketchpad, Utah
GM/IBM, Xerox, Microsoft
E&S, SGI, ATI, Adobe
Entertainment
Spacewar (MIT), Trek (Rochester)
Atari, Nintendo, SGI, Pixar
Internet
Arpanet, Aloha, Internet
Pup
DECnet, TCP/IP
LANs
Rings, Hubnet
Ethernet, Datakit, Autonet
LANs, switched Ethernet
Workstations
Lisp machine, Stanford
Xerox Alto
Xerox Star, Apollo, Sun
Graphical user interfaces
Engelbart / Rochester
Alto, Smalltalk
Star, Mac, Microsoft
VLSI design
Mead/Conway, Mosis
Berkeley, Caltech
many
RISC processors
Berkeley, Stanford
IBM 801
SUN, SGI, IBM, HP
to World Wide Web
from Internet
Relational databases
Berkeley, Wisconsin
IBM
Oracle, IBM, Sybase
Parallel databases
Tokyo, Wisconsin, UCLA
IBM, ICL
ICL, Teradata, Tandem
Data mining
Wisconsin, Stanford
IBM, Arbor
IRI, Arbor, Plato
Parallel computing
Illiac 4, CMU, Caltech, HPC
IBM, Intel
CM-5, Teradata, Cray T3D
RAID / disk servers
Berkeley
Striping /Datamesh, Petal
many
Portable communication
Berkeley, Purdue (CDMA)
Linkobit, Hughes
Qualcomm
World Wide Web
CERN, Illinois (Mosaic)
Alta Vista
Netscape, Yahoo
Speech recognition
CMU, SRI, MIT
Bell, IBM, Dragon
Dragon, IBM, L&H
Broadband last mile
Stanford, UCLA
Bellcore
Amati, Alcatel, Broadcom
University
Industry research
Products
IT research areas are ordered roughly according to when they became $1 billion industries.


Innovation
in INFORMATION TECHNOLOGY

2003


IT Sectors With Large Economic Impact
Motorola
AMD Intel
eBay Akamai Yahoo!
IBM
Electronic Arts
Qualcomm
HP Symantec
Juniper Facebook Twitter
VMware
HP
Adobe Autodesk
Nuance
Texas Instruments
Microsoft
XBox
nVidia
Apple
Cisco
Amazon
Oracle
nVidia Pixar
iRobot
iPhone
Dell
Google
iPod
Intuitive Surgical
Broadband & Mobile
Microprocessors
Personal Computing
Internet & Web
Cloud Computing
Enterprise Systems
Entertainment & Design
Robotics & Assistive Technologies
2010
2005
2000
1995
1990
1985
1980
1975
1970
1965
Digital Communications
Computer Architecture
Software Technologies
Networking
Parallel & Distributed Systems
Databases
Computer Graphics
AI & Robotics
Areas of Fundamental Research in IT
University
Industry R&D
Products
$1 Billion Market
$10 Billion Market


Continuing
Innovation
IN INFORMATION TECHNOLOGY

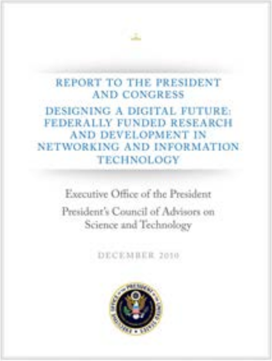
2012

- Key takeaways:
  - America is the world leader in information technology due to a rich interplay of government, academia, and industry
  - Every major market segment bears the clear stamp of Federal research investments
  - The path from research to major market segment is not linear: ideas and people flow in all directions
  - Unanticipated results are often as important as anticipated results
  - The interaction of research ideas multiplies their impact
  - Entirely appropriately, corporate R&D is very heavily tilted towards D: engineering the next release of a product, vs. a 5- 10- or 15-year horizon

# How should our competitiveness be defined?

- "At the time of the High-Performance Computing Act of 1991, the importance of high performance computing and communication (HPCC) to scientific discovery and national security was a major factor underlying the special attention given by Congress to NIT. Although HPCC continues to contribute in important ways to scientific discovery and national security, many other aspects of NIT have now risen to comparable levels of importance."

*N.B. This does <u>not</u> say that the importance of HPCC is decreasing! It simply notes that other aspects of the field have risen to comparable levels of importance, and must be weighed in assessing our competitiveness.*


REPORT TO THE PRESIDENT AND CONGRESS

DESIGNING A DIGITAL FUTURE: FEDERALLY FUNDED RESEARCH AND DEVELOPMENT IN NETWORKING AND INFORMATION TECHNOLOGY

Executive Office of the President

President's Council of Advisors on Science and Technology

DECEMBER 2010

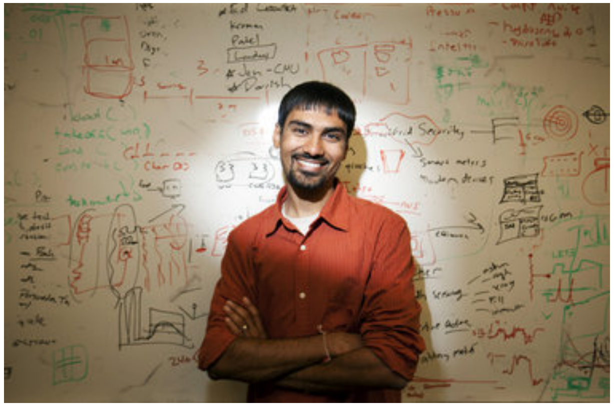
# The coming decade: Dramatic improvements in technology and algorithms enable “smart everything”

- A proliferation of sensors
  - Think about the sensors on your phone
- More generally, the creation of almost all information in digital form
  - It doesn’t need to be transcribed in order to be processed
- Dramatic cost reductions in storage
  - You can afford to keep all the data
- Dramatic increases in network bandwidth
  - You can move the data to where it’s needed

- Dramatic cost reductions and scalability improvements in computation
  - With Amazon Web Services, or Google App Engine, or Microsoft Azure, 1000 computers for 1 day costs the same as 1 computer for 1000 days
- Dramatic algorithmic breakthroughs
  - Machine learning, data mining – fundamental advances in computer science and statistics

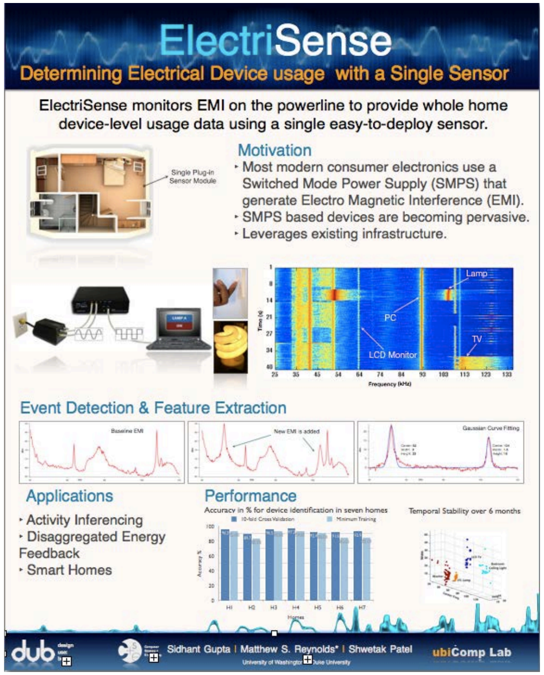
The "big data" revolution is what actually puts the "smarts" in "smart everything"

# Smart homes



Shwetak Patel,
University of Washington
2011 MacArthur Fellow

MACARTHUR
The John D. and Catherine T. MacArthur Foundation




belkin echo

# Smart cars

DARPA Grand Challenge


DARPA Urban Challenge



Google Self-Driving Car

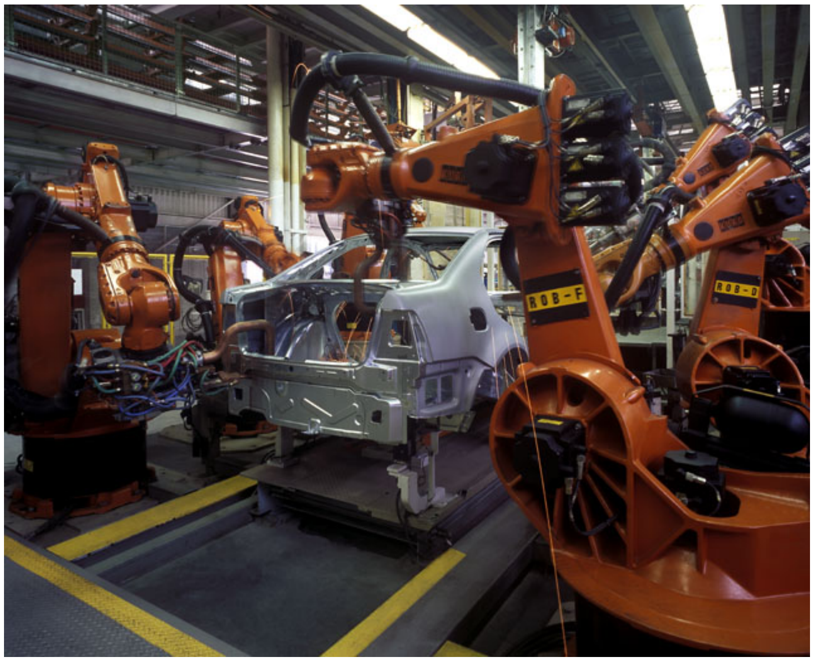
# Smart health

Larry Smarr –
"quantified self"

Evidence-based
medicine


P4 medicine

# Smart robots



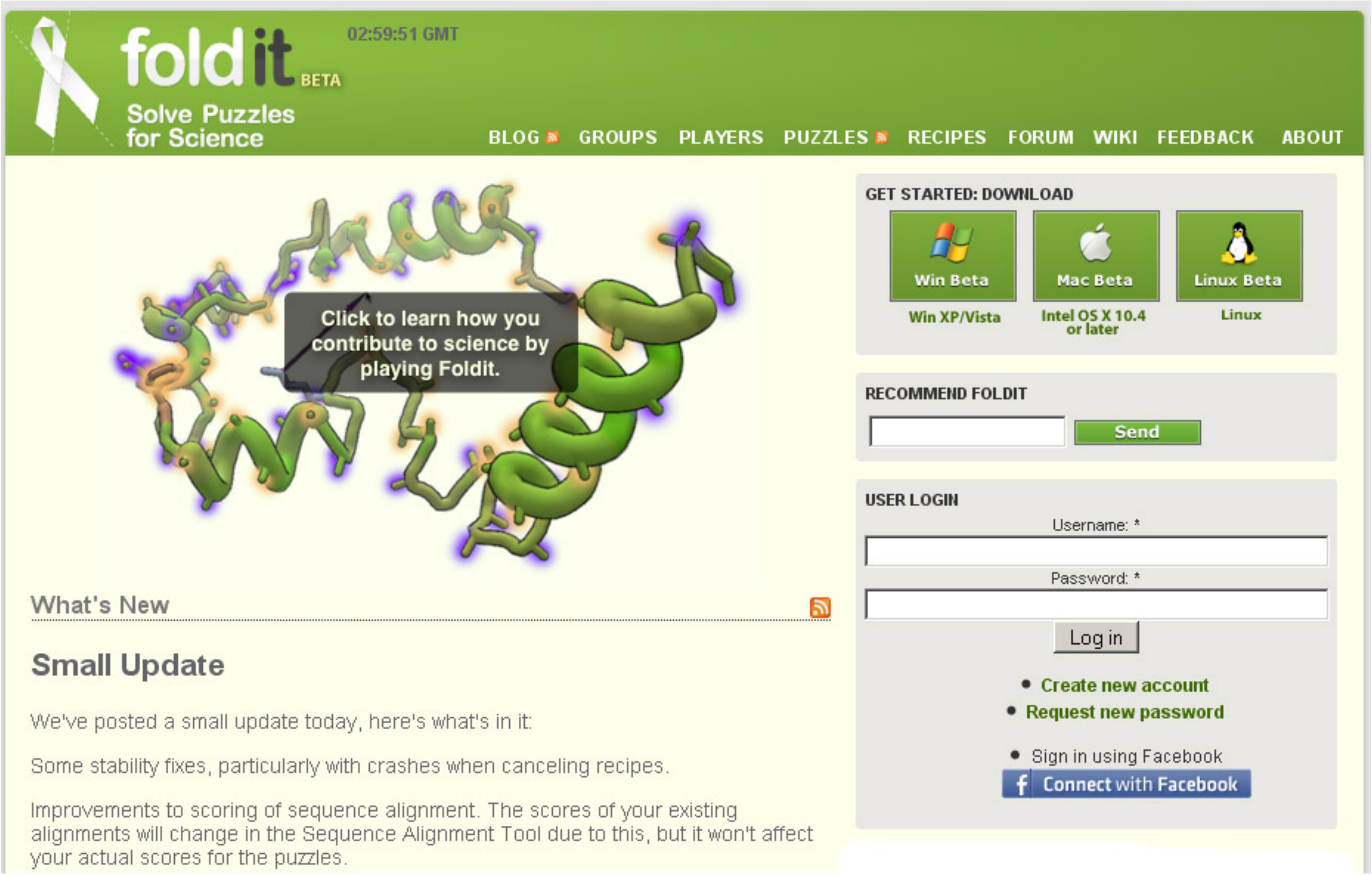

iRobot®

rethink robotics

# Smart crowds and human-computer systems

David Baker,
UW Biochemistry

Zoran Popovic,
UW Computer Science & Engineering

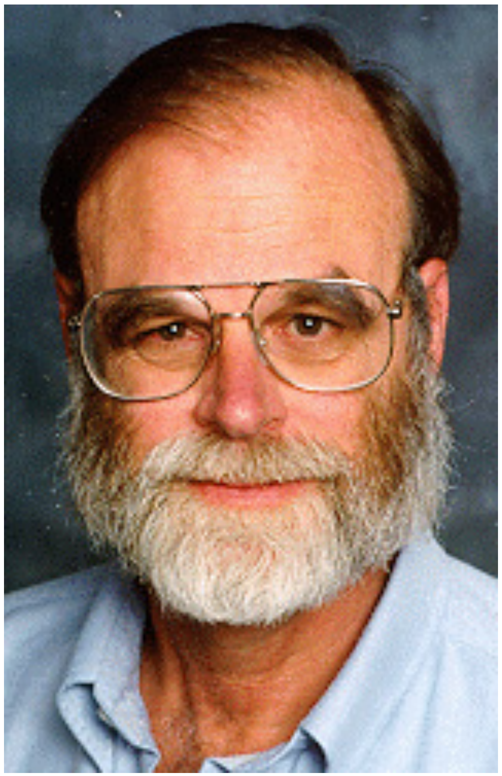


# Smart interaction


KINECT™

for XBOX 360

# Smart discovery (data-intensive discovery, or *eScience*)



Jim Gray,
Microsoft Research

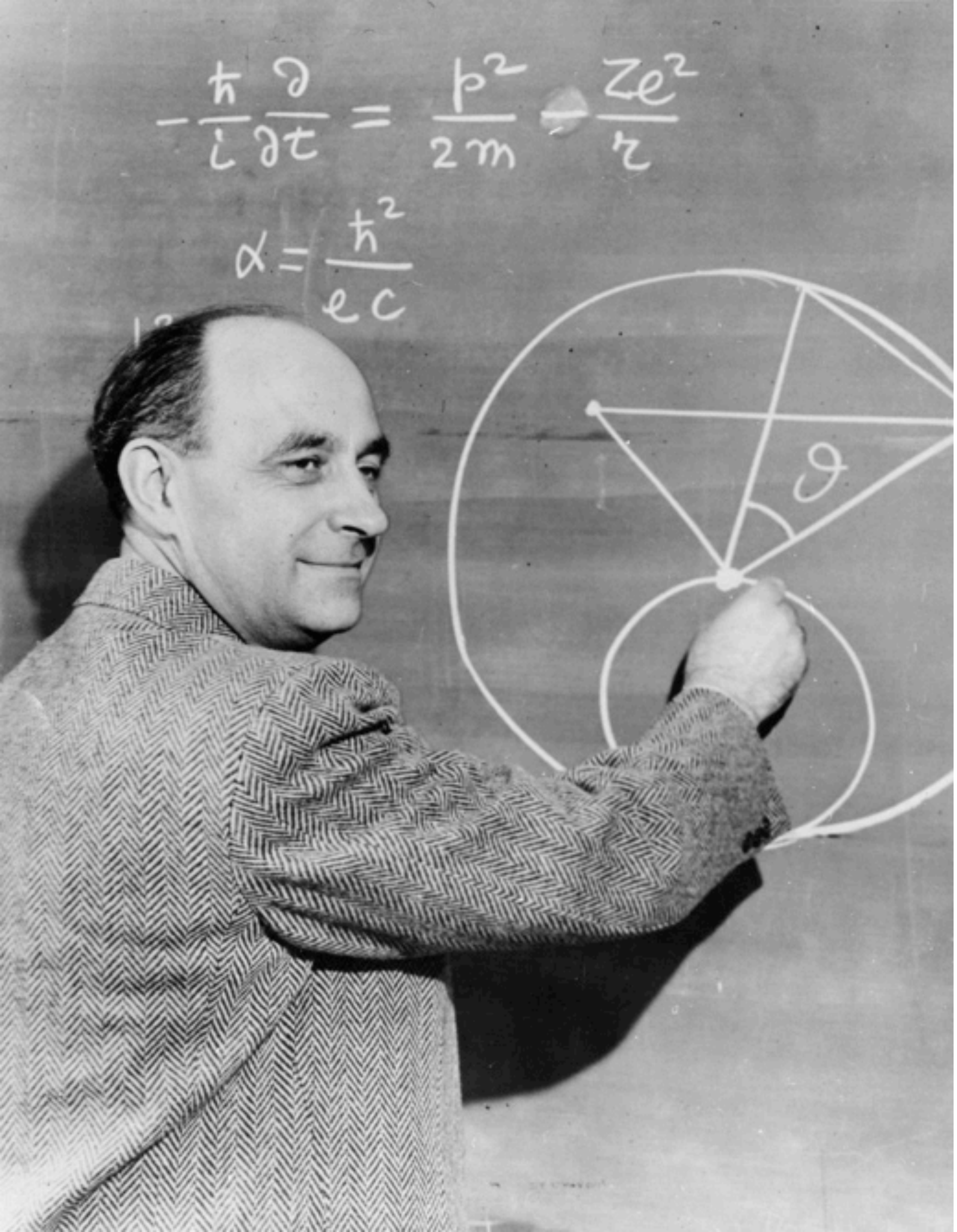

Transforming science (again!)


Observation
Experiment
Theory
Credit: John Delaney, University of Washington

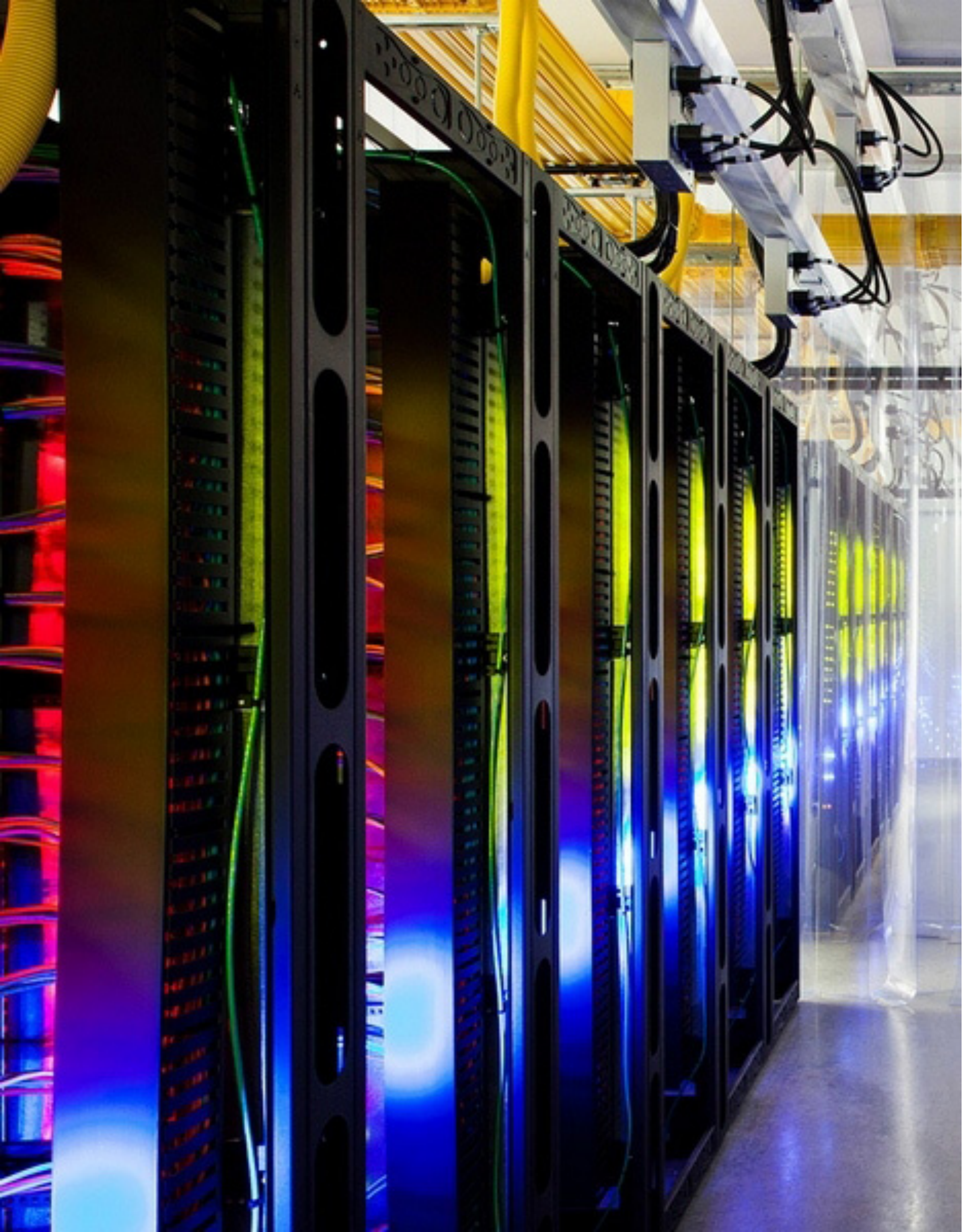
Observation

**Experiment**

Theory




Observation
Experiment
**Theory**



Observation

Experiment

Theory

**Computational Science**

Observation

Experiment

Theory

Computational Science

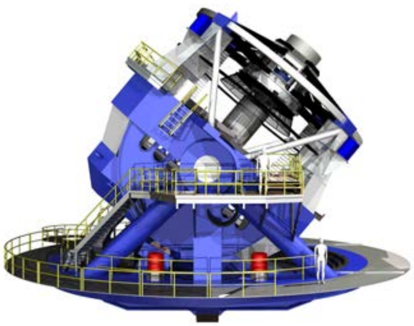
**eScience**

(*Augment*, not replace!)

SLOAN DIGITAL SKY SURVEY

# eScience is enabled by *data* more than by cycles

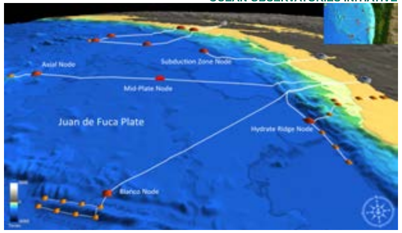
- Massive volumes of data from sensors and networks of sensors



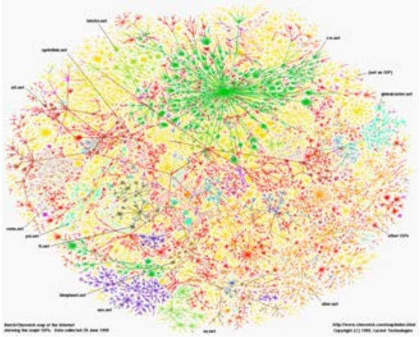
Astronomy: LSST

Physics: LHC

OCEAN OBSERVATORIES INITIATIVE




Oceanography: OOI



Sociology: The Web

Biology: Sequencing

Economics: POS terminals

Neuroscience: Hawkmoths

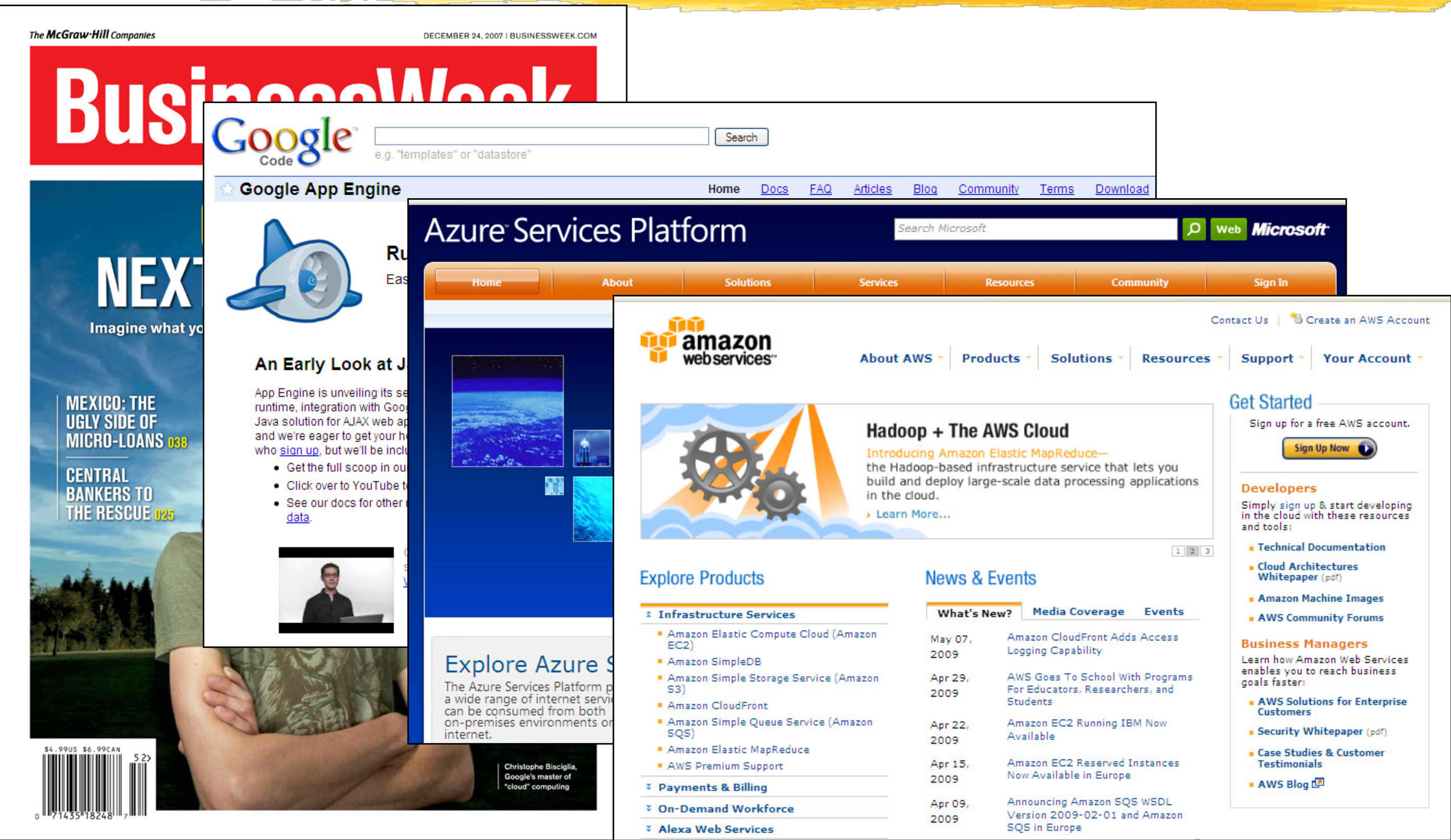
# eScience is about the *analysis* of data

- The semi-automated extraction of knowledge from massive volumes of data
  - There's simply too much of it – and it's too complex – to explore manually
- It's not just a matter of volume – it's "the 3 V's":
  - Volume
  - Velocity (rate)
  - Variety (dimensionality / complexity)
- It's not just a matter of data movement and data storage – it's about data *analysis* – "from data to knowledge to action"

# eScience utilizes a spectrum of computer science techniques and technologies

- Sensors and sensor networks
- Backbone networks
- Databases
- Data mining
- Machine learning
- Data visualization
- Cluster computing at enormous scale (the cloud)
- Collaboration and crowd sourcing

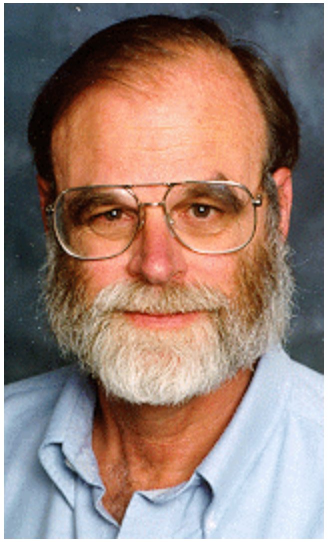
# eScience is married to the cloud: Scalable computing and storage for everyone



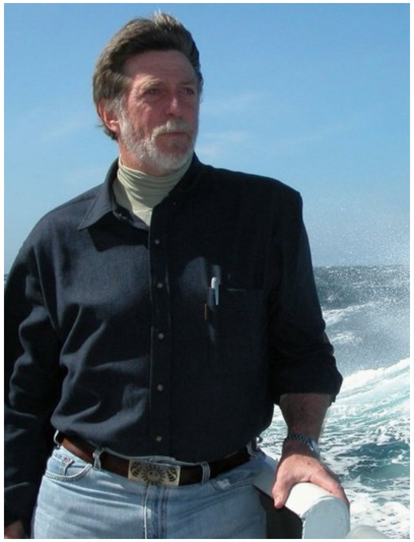
# eScience will be pervasive

- Simulation-oriented computational science has been transformational – and will continue to be of great importance – but it has *not* been pervasive
  - As an institution (e.g., a university), you didn't need to excel in order to remain competitive
- eScience capabilities must be broadly available in any institution
  - If not, the institution will simply cease to be competitive

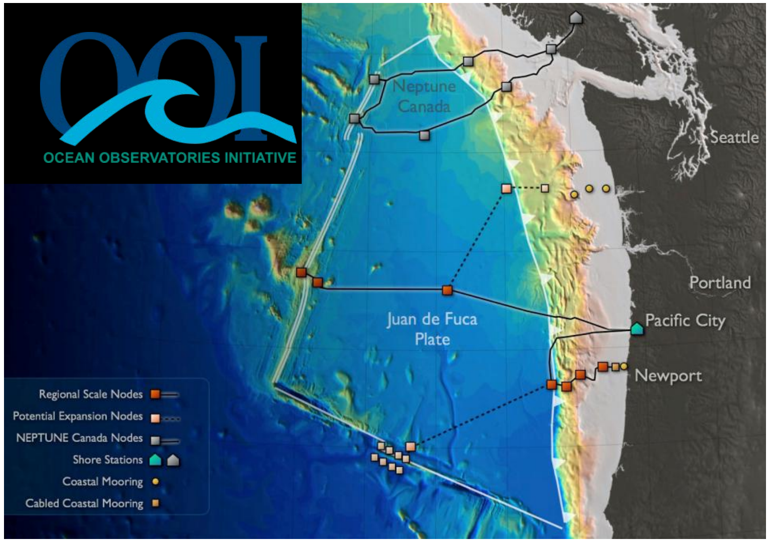
# My personal story, and the story of the UW eScience Institute



Early 1980s

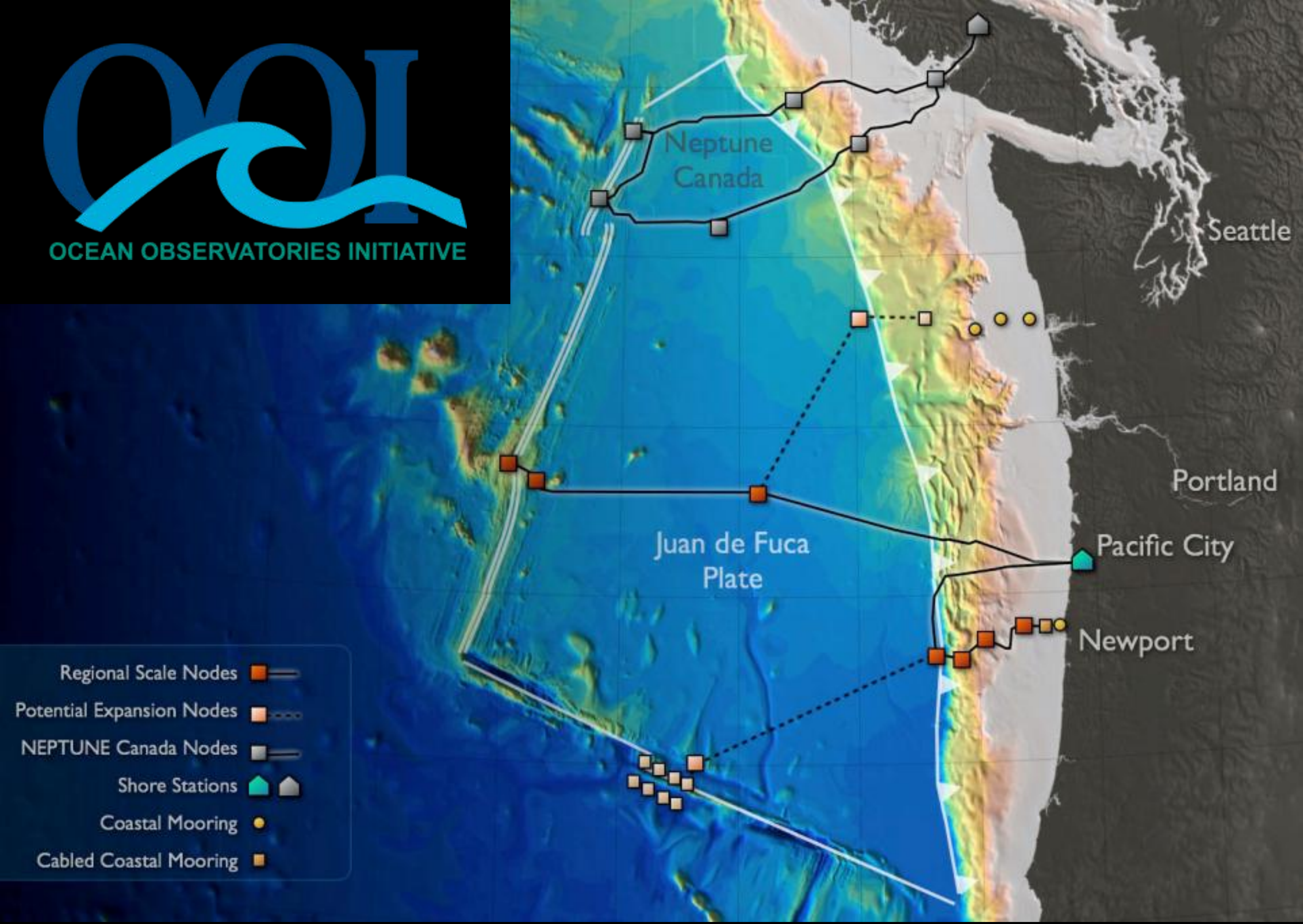


Late 1990s




Credit: John Delaney, University of Washington

OCEAN OBSERVATORIES INITIATIVE




Credit: John Delaney, University of Washington

Mark Emmert

2004

Ed Lazowska
Computer Science & Engineering

Tom Daniel
Biology

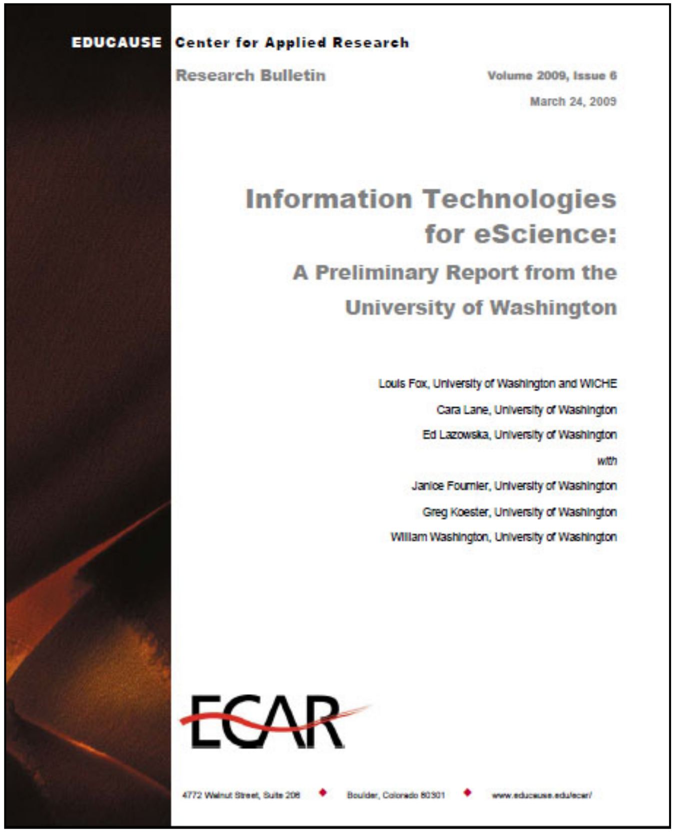
Werner Stuetzle
Statistics

# UW eScience Institute

- *"All across our campus, the process of discovery will increasingly rely on researchers' ability to extract knowledge from vast amounts of data... In order to remain at the forefront, UW must be a leader in advancing these techniques and technologies, and in making [them] accessible to researchers in the broadest imaginable range of fields."*
- In other words:
  - Data-intensive discovery will be ubiquitous
  - We must be a leader in inventing the capabilities
  - We must be a leader in translational activities – in putting these capabilities to work
  - It's about *intellectual infrastructure* (human capital) and *software infrastructure* (shared tools and services – digital capital)

# This was not as broadly obvious in 2005 as it is today

- But we asked UW's leading faculty – across all ages and fields, regardless of "label" – and they confirmed this view of the future
  - *From the start, this effort has been bottom-up, needs-based, driven by the scientists*


EDUCAUSE Center for Applied Research

Research Bulletin

Volume 2009, Issue 6

March 24, 2009

Information Technologies for eScience:

A Preliminary Report from the University of Washington

Louis Fox, University of Washington and WICHE

Cara Lane, University of Washington

Ed Lazowska, University of Washington

with

Janice Fournier, University of Washington

Greg Koester, University of Washington

William Washington, University of Washington

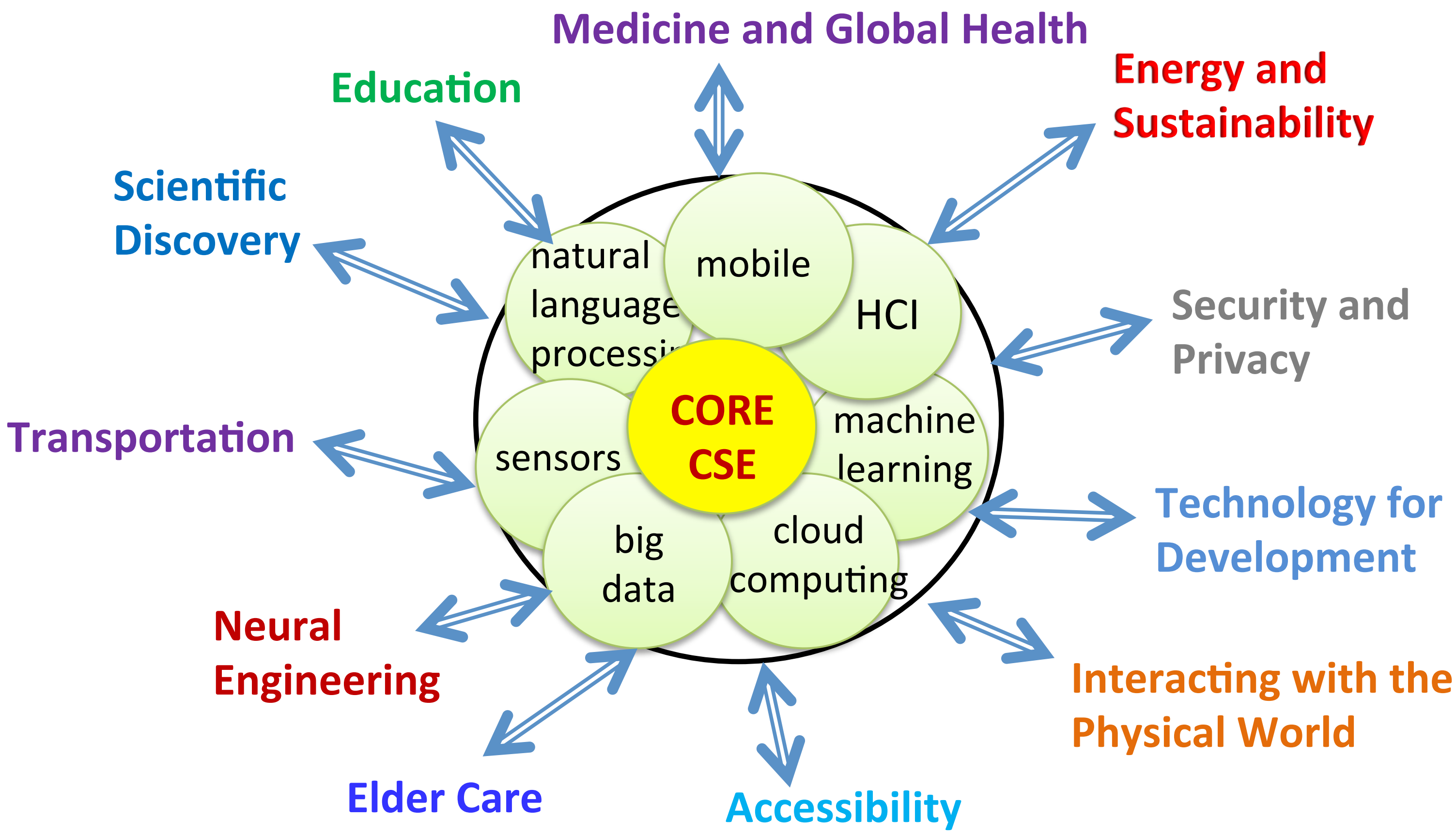
ECAR

4772 Walnut Street, Suite 206 ♦ Boulder, Colorado 80301 ♦ www.educause.edu/ecar/

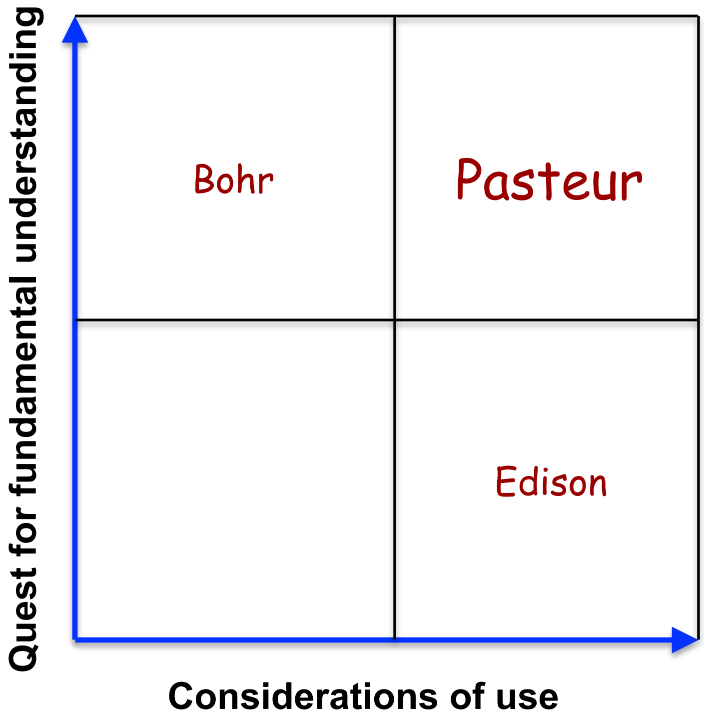
# We're at the dawn of a revolutionary new era of discovery and of learning

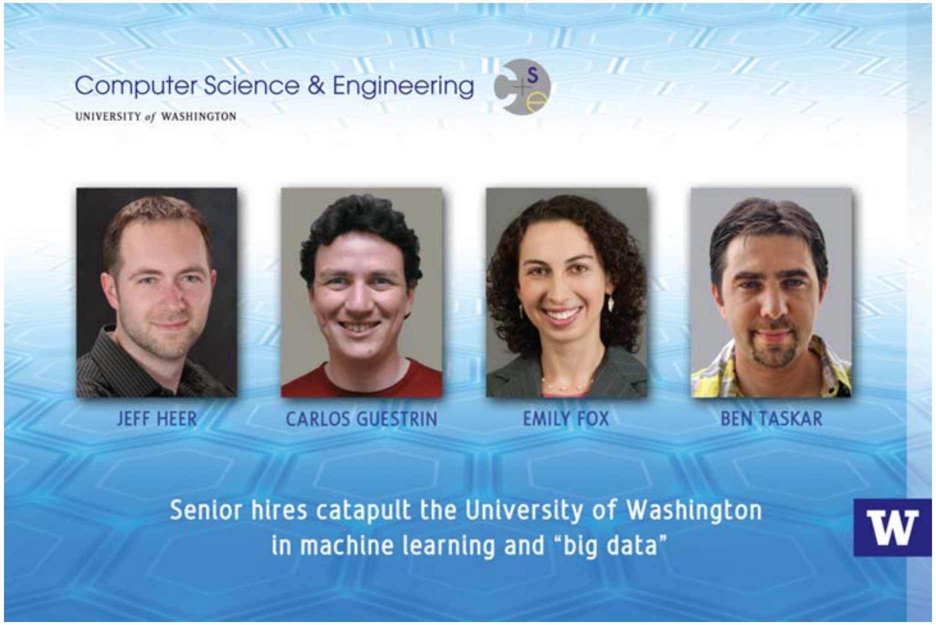
# Implications for academia

- A modern view of computer science

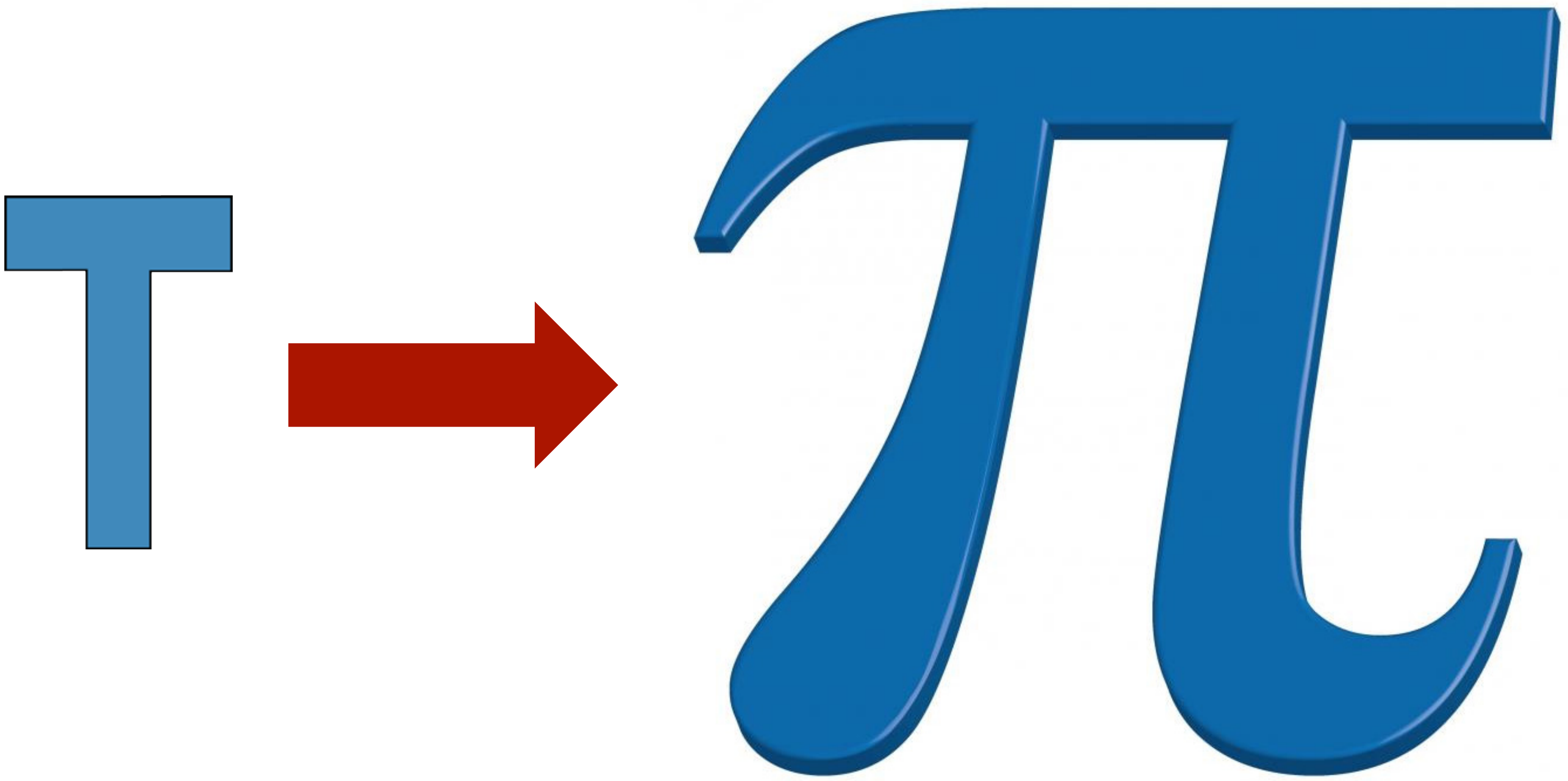


- On the methodology side, seek faculty in "Pasteur's quadrant"



- Strive to educate students who are "Pi-shaped"



Credit: Alex Szalay, Johns Hopkins University

- Resurrect the water cooler!  (Data is today's unifying force!)

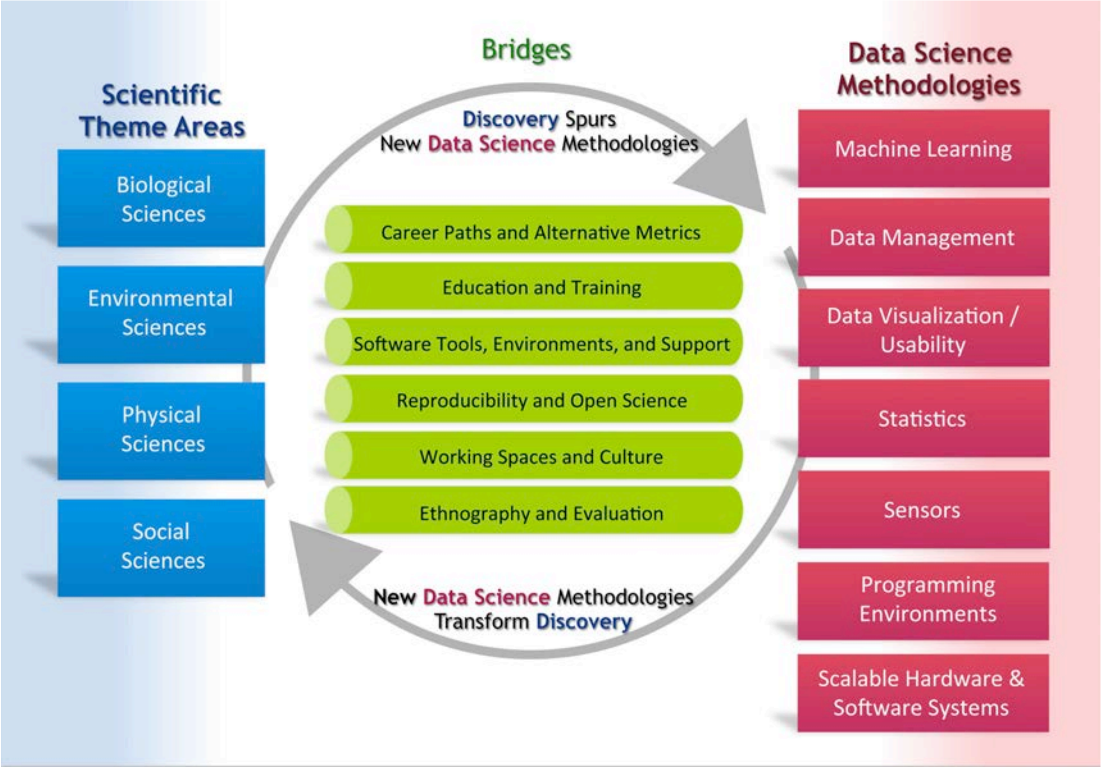


# Implications for research policy

- NSF has a unique role in driving advances in computer science
  - Computer Science does not have an NIH or a Department of Energy
  - NSF provides 75% of Federal support for academic computer science research
- Other fields are becoming *information* fields, not just computational fields
  - The *intellectual approaches* of computer science are as important to advances as is cyberinfrastructure
  - *New approaches* will enable *new discoveries*

- Meeting evolving cyberinfrastructure needs requires *research*, not merely procurement
  - This is true for HPC … and for data-intensive discovery … and for cyber-enabled advances in education and assessment
- Meeting evolving cyberinfrastructure needs requires investment in *intellectual* as well as physical infrastructure

- Advancing data-intensive discovery requires broad-based programs that strive to create a "virtuous cycle" – and that drive *institutional change*



- Nationally and institutionally, there are various policies that distort behavior
  - One example: Appropriate use of cloud resources is discouraged by
    - Indirect cost on services
    - MRI viewed as a pot separate from Directorates/Divisions
    - Institutional subsidies (power, cooling, space)

# Implications for K-12 education
# Computer Science in K-12, 1983

## A Nation At Risk

Our Nation is at risk. Our once unchallenged preeminence in commerce, industry, science, and technological innovation is being overtaken by competitors throughout the world. This report is concerned with only one of the many causes and dimensions of the problem, but it is the one that undergirds American prosperity, security, and civility.

If an unfriendly foreign power had attempted to impose on America the mediocre educational performance that exists today, we might well have viewed it as an act of war. As it stands, we have allowed this to happen to ourselves.

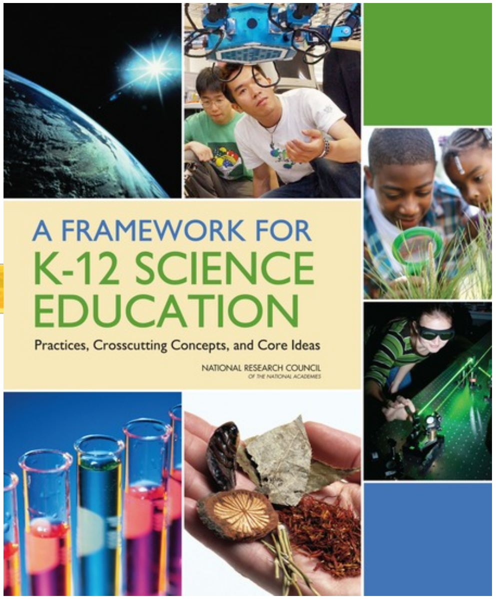
## Recommendation A: Content

**We recommend** *that State and local high school graduation requirements be strengthened and that,* at a minimum, all *students seeking a diploma be required to lay the foundations in the Five New Basics by taking the following curriculum during their 4 years of high school: (a) 4 years of English; (b) 3 years of mathematics; (c) 3 years of science; (d) 3 years of social studies; and (e) one-half year of computer science.*

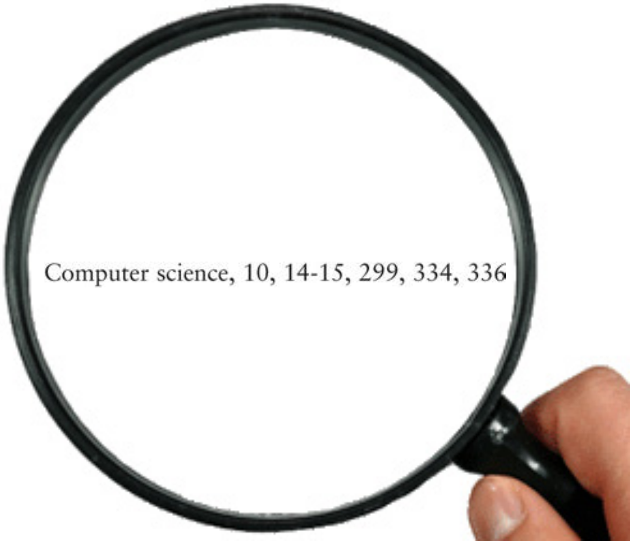

IBM PC XT
4.77 MHz 8088
128 KB RAM
PC DOS 2.0

# Computer Science in K-12, 2013






401 page report
15 page index



NEXT GENERATION SCIENCE STANDARDS

| Elementary (K–5) | | |
|---|---|---|
| Storylines: K–2 3–5 PDFs: K 1 2 3 4 5 | | |
| **K.** Forces and Interactions: Pushes and Pulls | **2.** Interdependent Relationships in Ecosystems | **4.** Waves |
| **K.** Interdependent Relationships in Ecosystems: Animals, Plants, and Their Environment | **2.** Earth's Systems: Processes that Shape the Earth | **4.** Structure, Function, and Information Processing |
| **K.** Weather and Climate | **K-2.** Engineering Design | **4.** Earth's Systems: Processes that Shape the Earth |
| **1.** Waves: Light and Sound | **3.** Forces and Interactions | **5.** Structure and Properties of Matter |
| **1.** Structure, Function and Information Processing | **3.** Interdependent Relationships in Ecosystems | **5.** Matter and Energy in Organisms and Ecosystems |
| **1.** Space Systems: Patterns and Cycles | **3.** Inheritance and Variation of Traits | **5.** Earth's Systems |
| **2.** Structure and Properties of Matter | **3.** Weather and Climate | **5.** Space Systems: Stars and the Solar System |
| | **4.** Energy | **3-5.** Engineering Design |

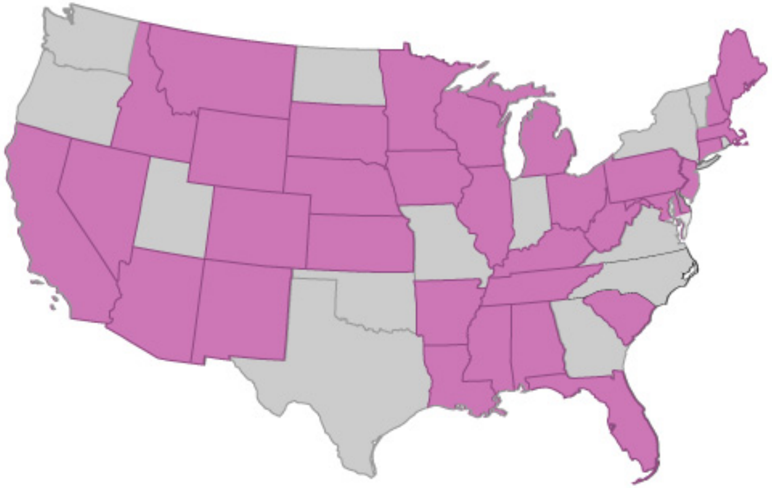
| PS: Physical Sciences | |
|---|---|
| Middle School (6–8) Storyline PDF | High School (9–12) Storyline PDF |
| **MS.** Structure and Properties of Matter | **HS.** Structure and Properties of Matter |
| **MS.** Chemical Reactions | **HS.** Chemical Reactions |
| **MS.** Forces and Interactions | **HS.** Forces and Interactions |
| **MS.** Energy | **HS.** Energy |
| **MS.** Waves and Electromagnetic Radiation | **HS.** Waves and Electromagnetic Radiation |

| LS: Life Sciences | |
|---|---|
| Middle School (6–8) Storyline PDF | High School (9–12) Storyline PDF |
| **MS.** Structure, Function, and Information Processing | **HS.** Structure and Function |
| **MS.** Matter and Energy in Organisms and Ecosystems | **HS.** Matter and Energy in Organisms and Ecosystems |
| **MS.** Interdependent Relationships in Ecosystems | **HS.** Interdependent Relationships in Ecosystems |
| **MS.** Growth, Development, and Reproduction of Organisms | **HS.** Inheritance and Variation of Traits |
| **MS.** Natural Selection and Adaptations | **HS.** Natural Selection and Evolution |

| ESS: Earth and Space Sciences | |
|---|---|
| Middle School (6–8) Storyline PDF | High School (9–12) Storyline PDF |
| **MS.** Space Systems | **HS.** Space Systems |
| **MS.** History of Earth | **HS.** History of Earth |
| **MS.** Earth's Systems | **HS.** Earth's Systems |
| **MS.** Weather and Climate | **HS.** Weather and Climate |
| **MS.** Human Impacts | **HS.** Human Sustainability |

| ETS: Engineering, Technology, and Applications of Science | |
|---|---|
| Middle School (6–8) Storyline PDF | High School (9–12) Storyline PDF |
| **MS.** Engineering Design | **HS.** Engineering Design |

- In 9 out of 10 high schools nationwide, computer science is not offered
- In 36 of the 50 states, computer science does not count towards the math or science graduation requirement



Yet computer science – “computational thinking” – is a key capability for just about every 21st century endeavor

# Is this a great time, or what?!?!


http://lazowska.cs.washington.edu/NSB.pdf