eScience-NIAC Partnership

Ed Lazowska

Bill & Melinda Gates Chair in
Computer Science & Engineering

Founding Director, eScience Institute

University of Washington

NIAC Day at UW

October 2013

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

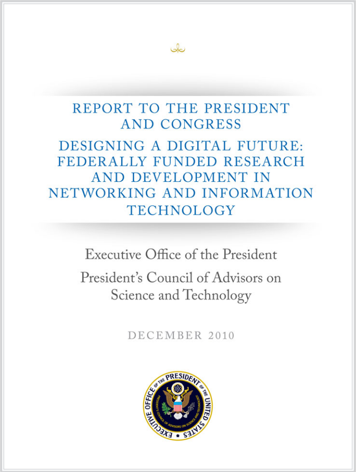
This morning …

[Adaptation of a September presentation to the National Science Board on "The Future of Advanced Cyberinfrastructure for Science and Engineering Research and Education"]

- Why must America remain the world leader in computing?
- 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"

Why must America remain the world leader in computing?

- "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

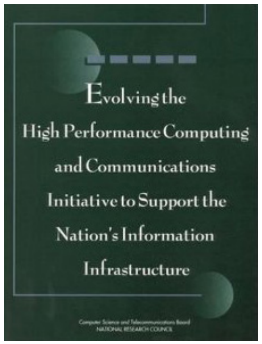
DESIGNING A DIGITAL FUTURE: FEDERALLY FUNDED RESEARCH AND DEVELOPMENT IN NETWORKING AND INFORMATION TECHNOLOGY

Executive Office of the President

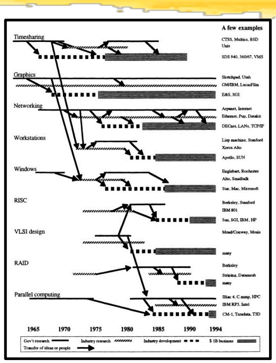
President's Council of Advisors on Science and Technology

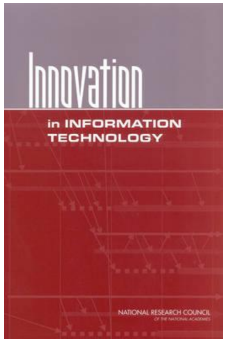
DECEMBER 2010

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

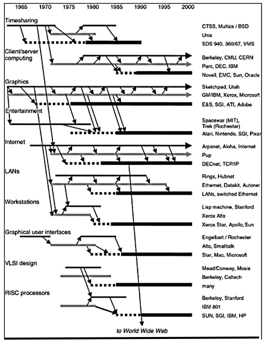


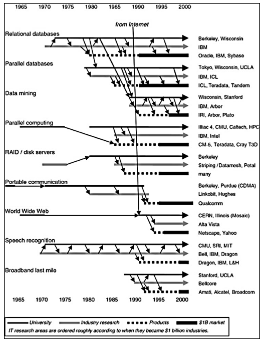
1995



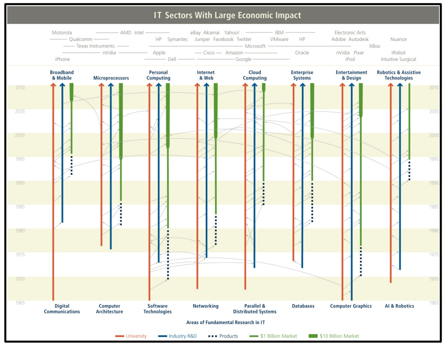


2003





2012

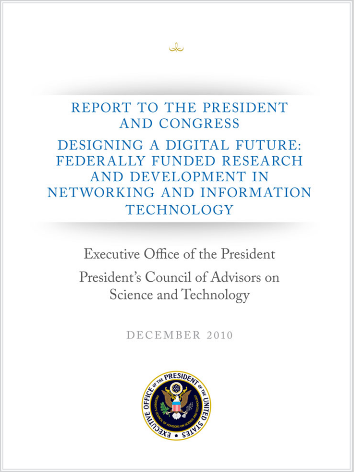


- Key takeaways:
  - America is the world leader in information technology due to a rich interplay of government, academia, and industry
  - Every major market segment bears the clear stamp of Federal research investments
  - The path from early research to major market segment is not linear: ideas and people flow in all directions
  - That path typically requires 15 years to traverse
  - 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 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

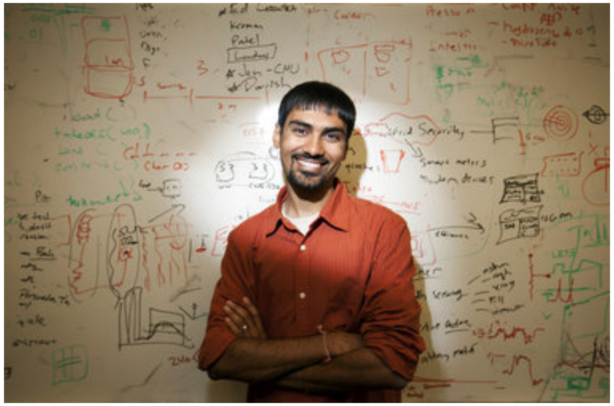
The coming decade: Dramatic improvements in technology and algorithms enable "smart everything"

- 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

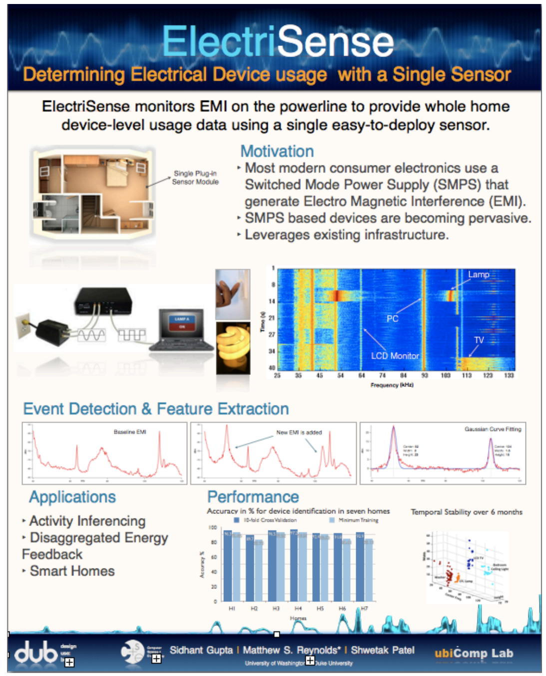
The "big data" revolution is what actually puts the "smarts" in "smart everything"

Smart homes (the leaf nodes of the smart grid)



Shwetak Patel,
University of Washington
2011 MacArthur Fellow

MACARTHUR
The John D. and Catherine T. MacArthur Foundation

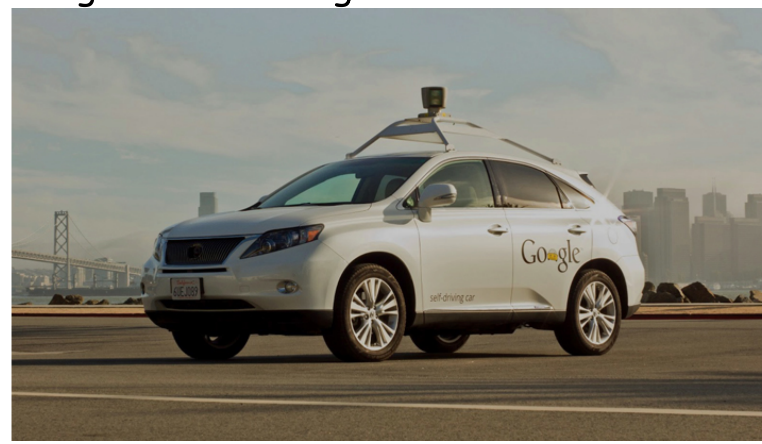


belkin echo

Smart cars

DARPA Grand Challenge

Google Self-Driving Car



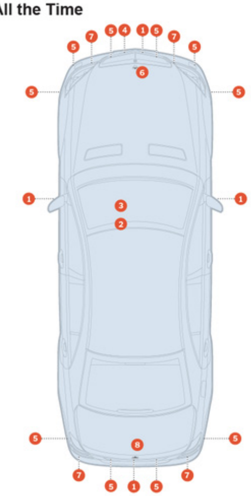
DARPA Urban Challenge

2014 Mercedes-Benz S-Class

Looking in All Directions, All the Time

The systems that provide assistance for drivers of the 2014 Mercedes-Benz S-Class depend on cameras, radar units and ultrasonic sensors to keep an eye on other vehicles, pedestrians, animals and even the road surface. With a 360-degree view of the car's environment, little is missed. Capable of steering, braking, throttle control and even decision-making, the automaker's Intelligent Drive technology can operate the vehicle without driver input for a limited amount of time and in certain prescribed circumstances.

1 Cameras
2 Stereo multipurpose camera
3 Near-infrared camera
4 Far-infrared camera
5 Ultrasonic sensors
6 Long-range radar
7 Short-range radar
8 Multimode radar



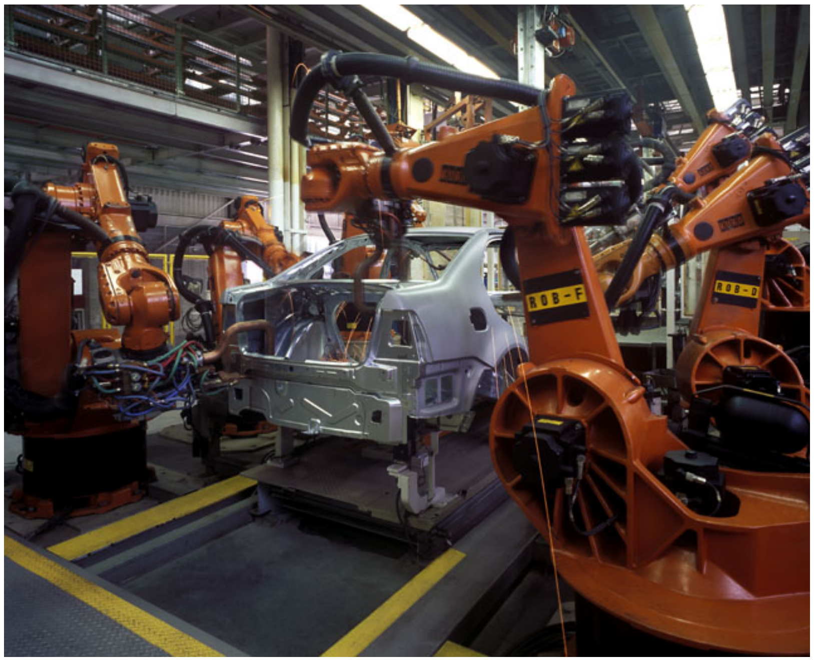
Smart health

Larry Smarr – "quantified self"

Evidence-based medicine

P4 medicine

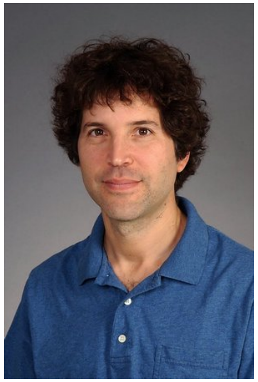
Smart robots



iRobot

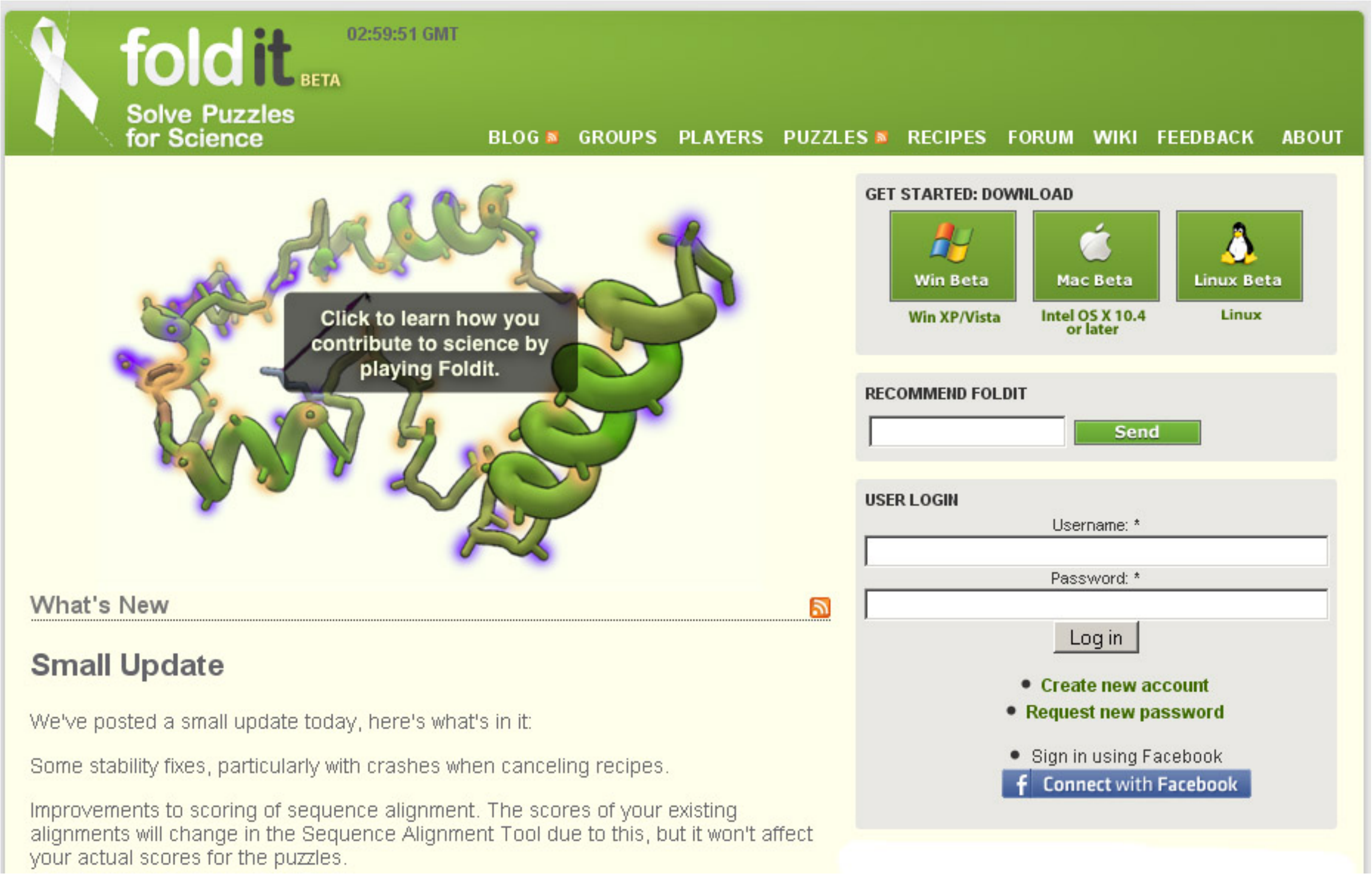
rethink robotics

Smart crowds and human-computer systems



David Baker,
UW Biochemistry

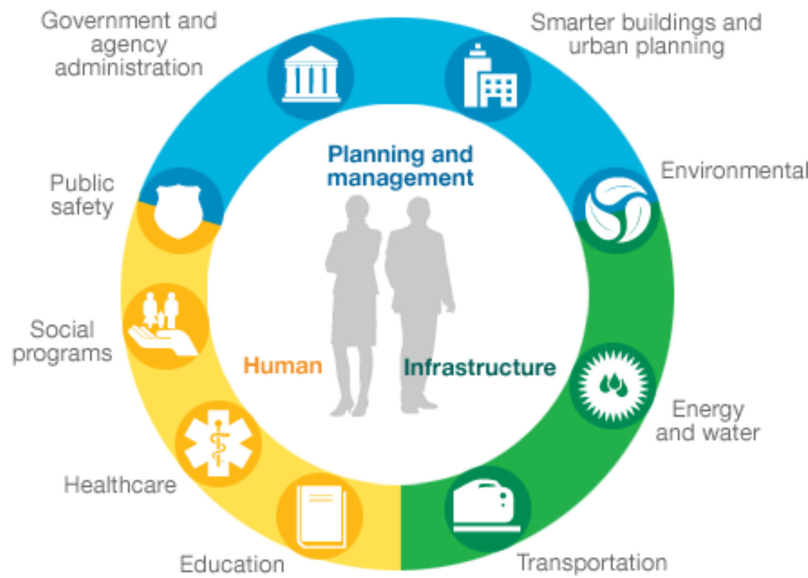
Zoran Popovic,
UW Computer Science & Engineering



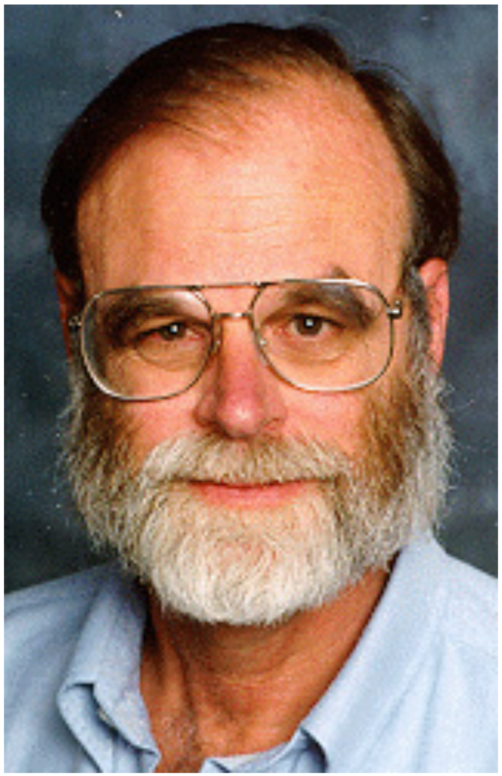
Smart interaction

KINECT™
for XBOX 360

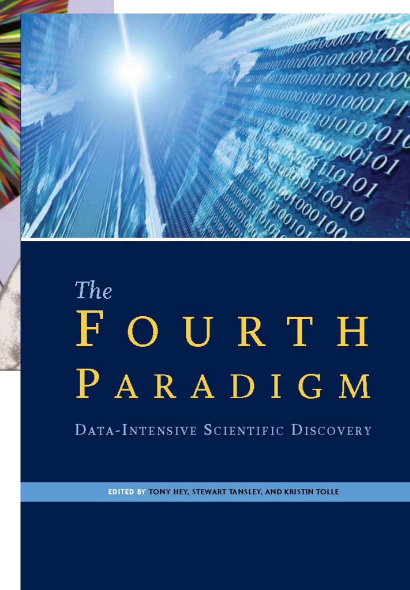
Smart cities



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



Jim Gray,
Microsoft Research



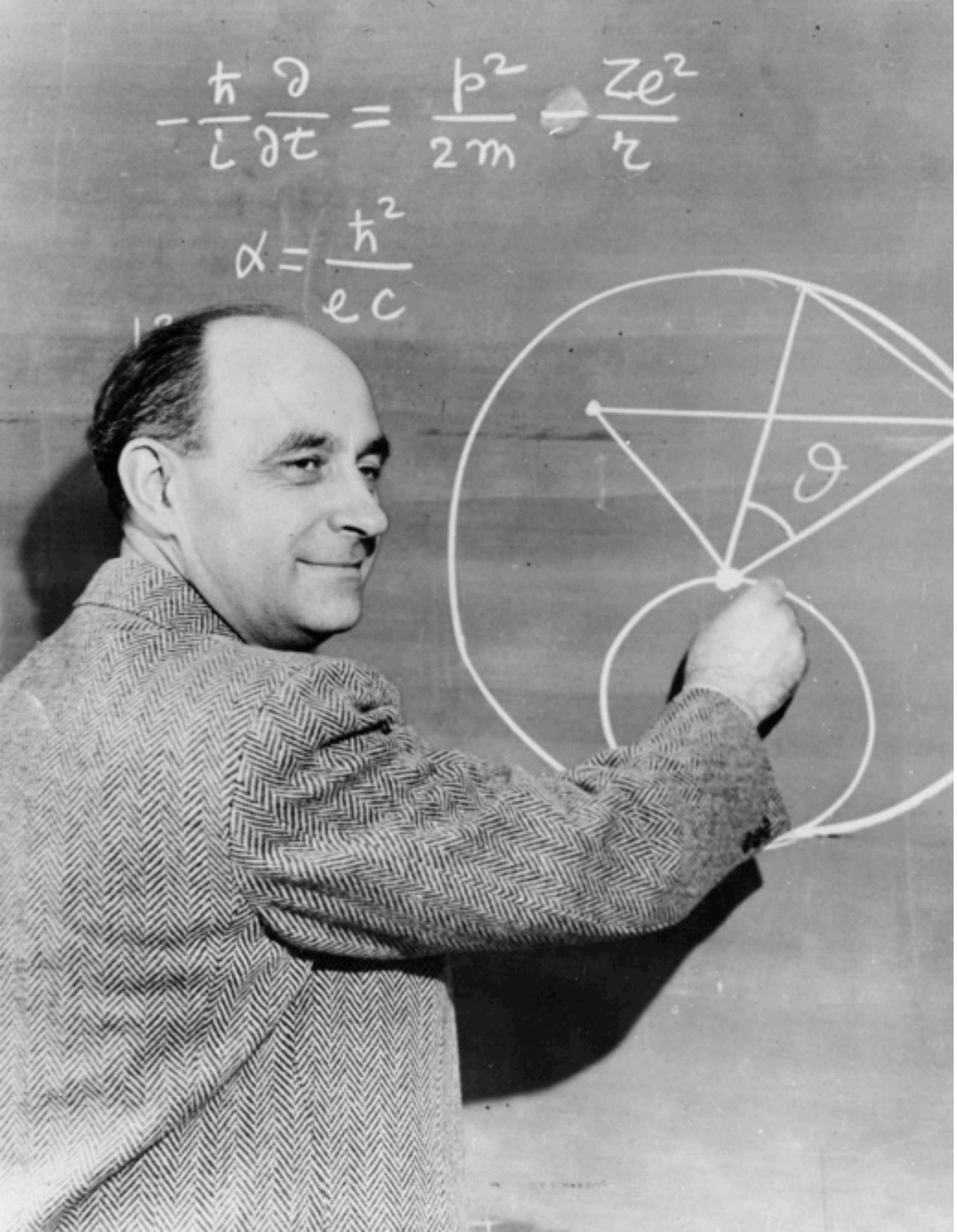
Transforming science (again!)


Observation
Experiment
Theory
Credit: John Delaney, University of Washington

Observation

Experiment

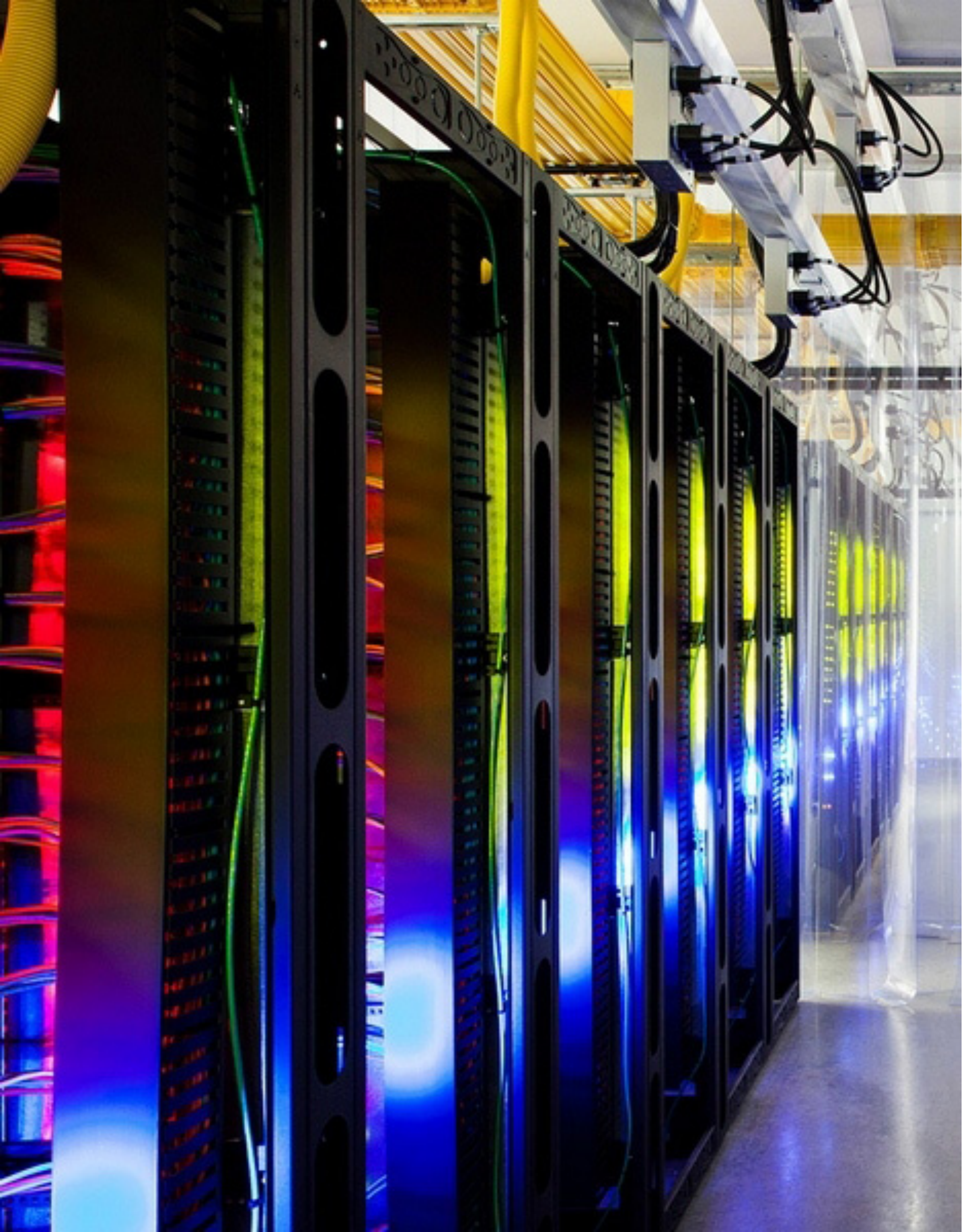
Theory



Observation

Experiment

Theory



Observation

Experiment

Theory

Computational Science

Observation

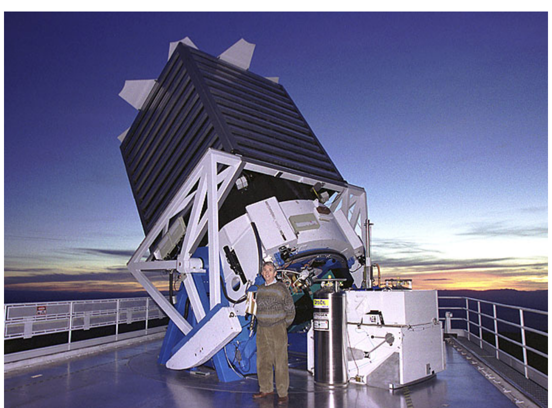
Experiment

Theory

Computational Science

eScience

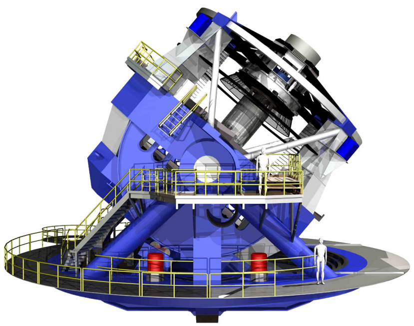
(*Augment*, not replace!)



SLOAN DIGITAL SKY SURVEY

eScience is enabled by *data* more than by cycles

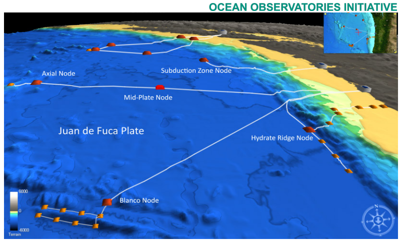
- Massive volumes of data from sensors and networks of sensors



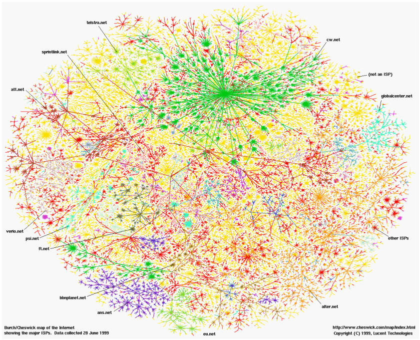
Astronomy: LSST

Physics: LHC

OOI
OCEAN OBSERVATORIES INITIATIVE



Oceanography: OOI



Sociology: The Web

Biology: Sequencing

Economics: POS terminals

Neuroscience: Hawkmoths

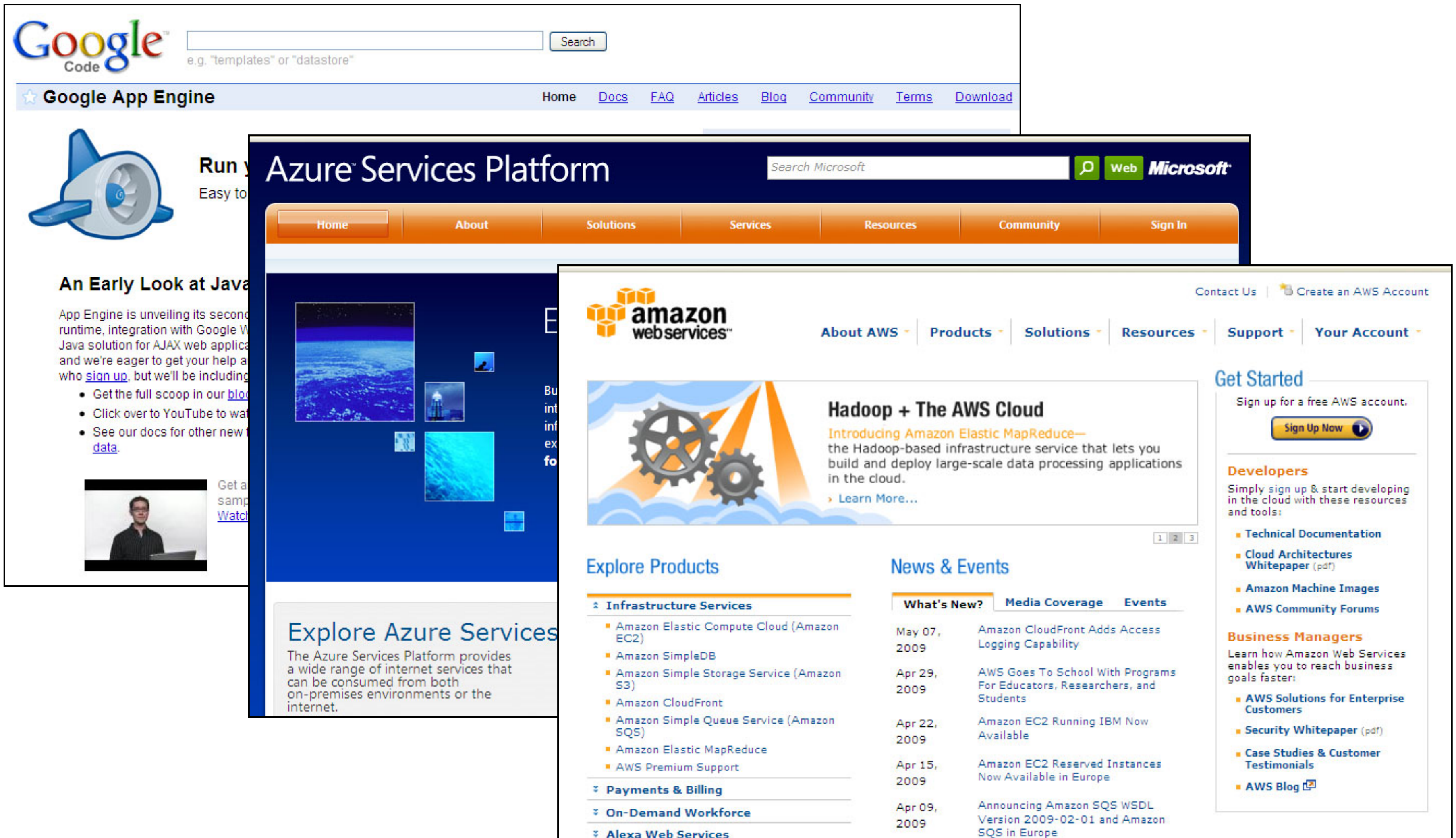
eScience is about the *analysis* of data

- 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

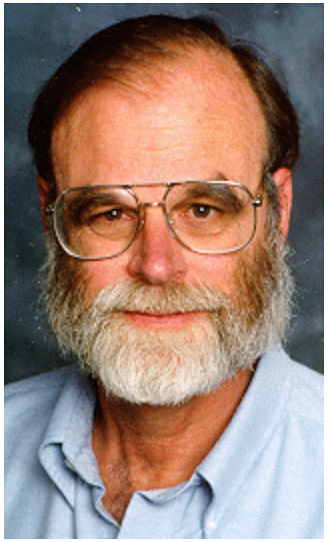
eScience is married to the cloud: Scalable computing and storage for everyone



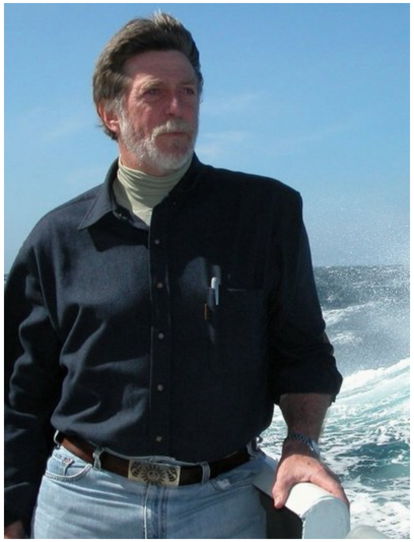
eScience will be *pervasive*

- 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

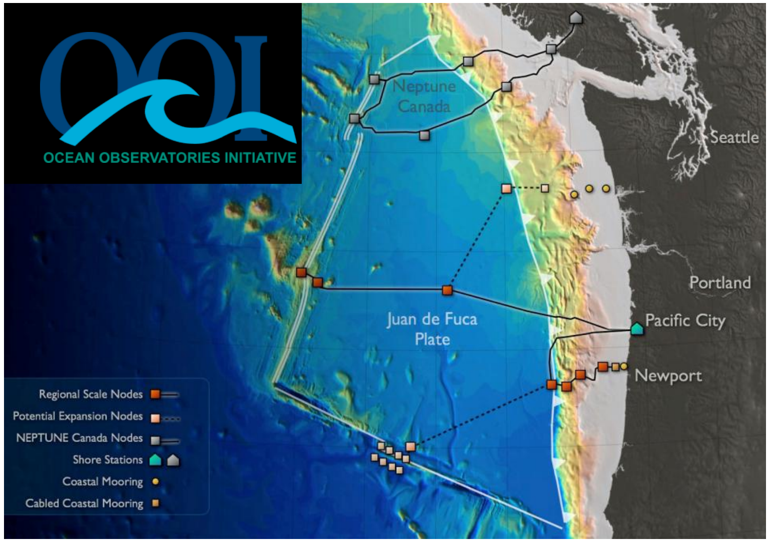
My personal story, and the story of the UW eScience Institute



Early 1980s

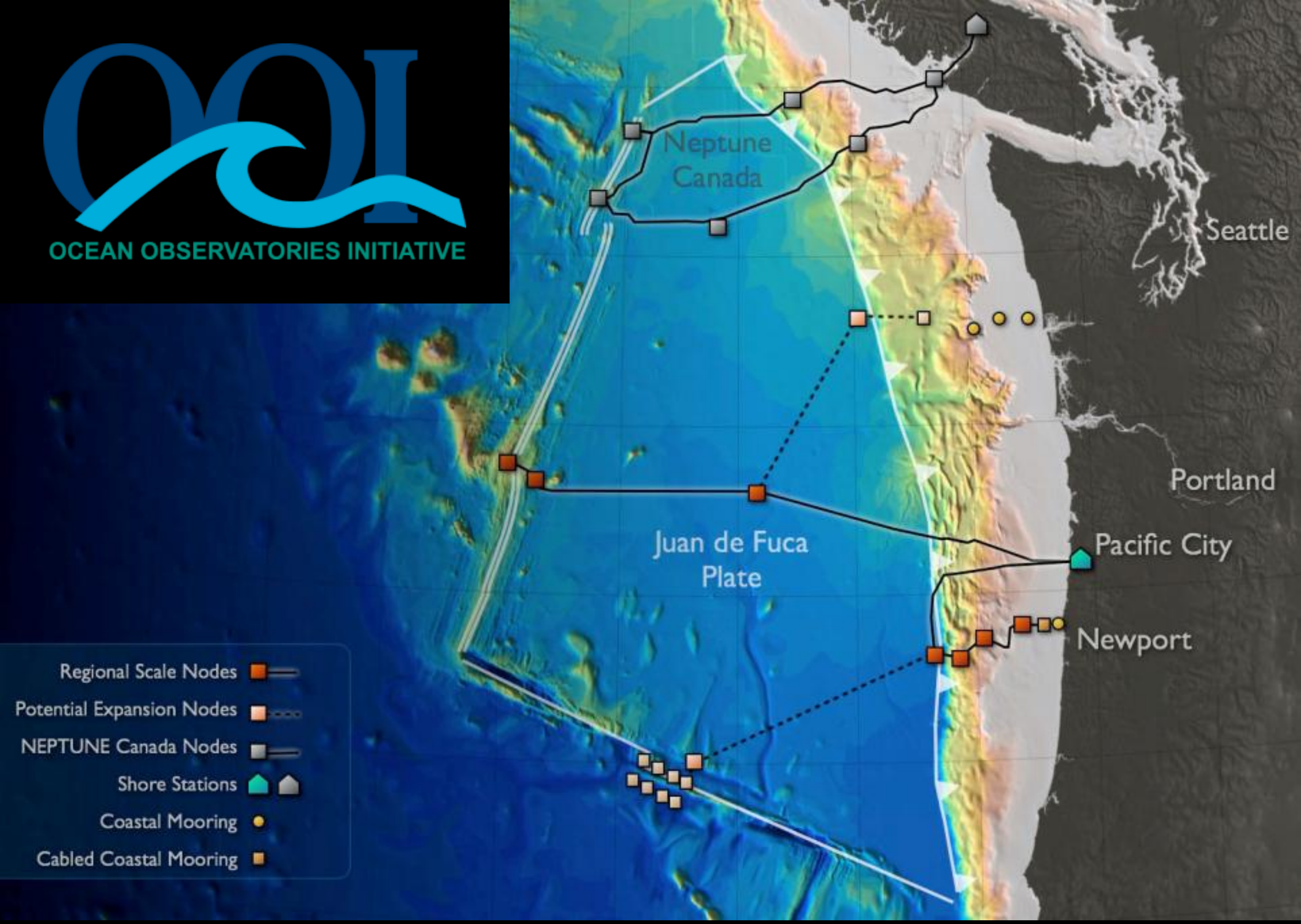


Late 1990s




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

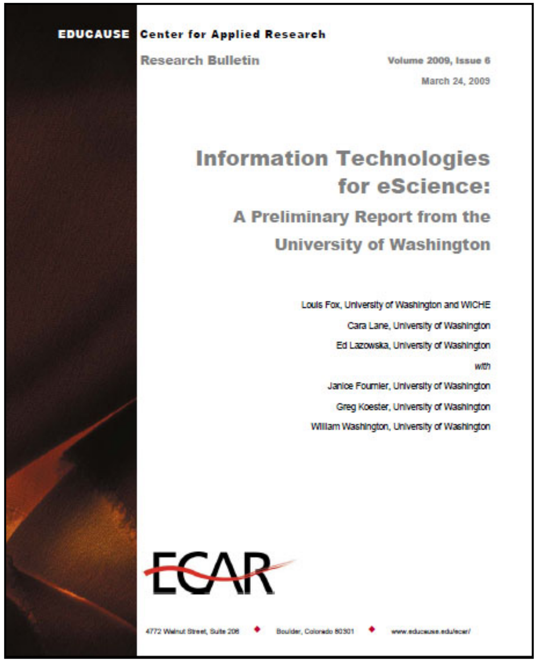
UW eScience Institute: Concept

- *"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)

eScience Institute

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

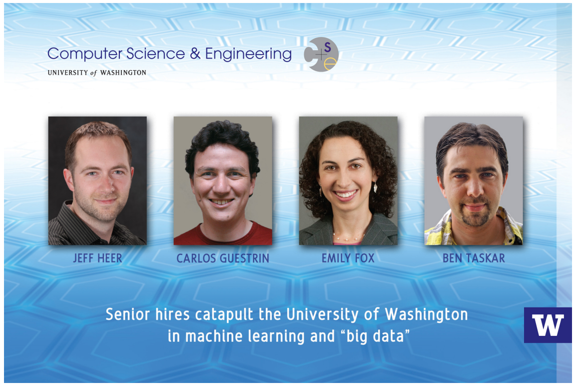
ECAR

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

eScience Institute

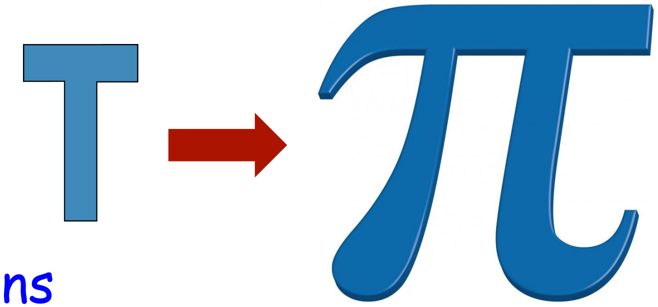
UW eScience Institute: Today

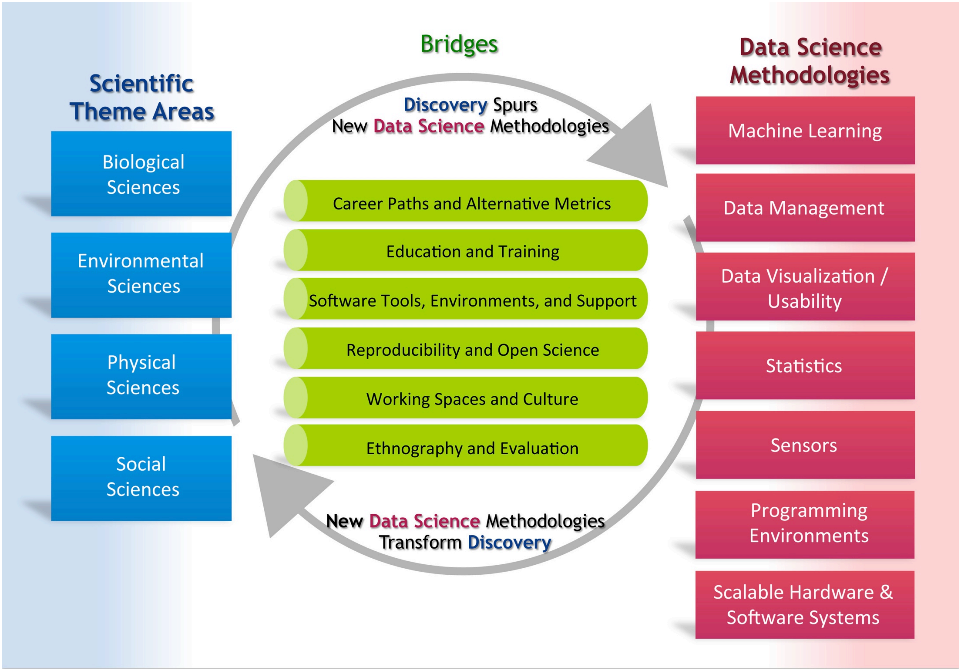
- More than 50 affiliated faculty
- 11-member Steering Committee
- Expanding cadre of Research Scientists
- Signature hires in data science methodology
  - Additional slots committed by the Provost



- Multiple high-impact engagements between methodology researchers (computer science, statistics, applied mathematics) and colleagues in astronomy, biology, geography, oceanography, sociology, …
- Large number of cross-disciplinary research grants and contracts
- Vibrant seminar series
- Significant partnerships with Amazon.com, Google, Microsoft, Tableau, …

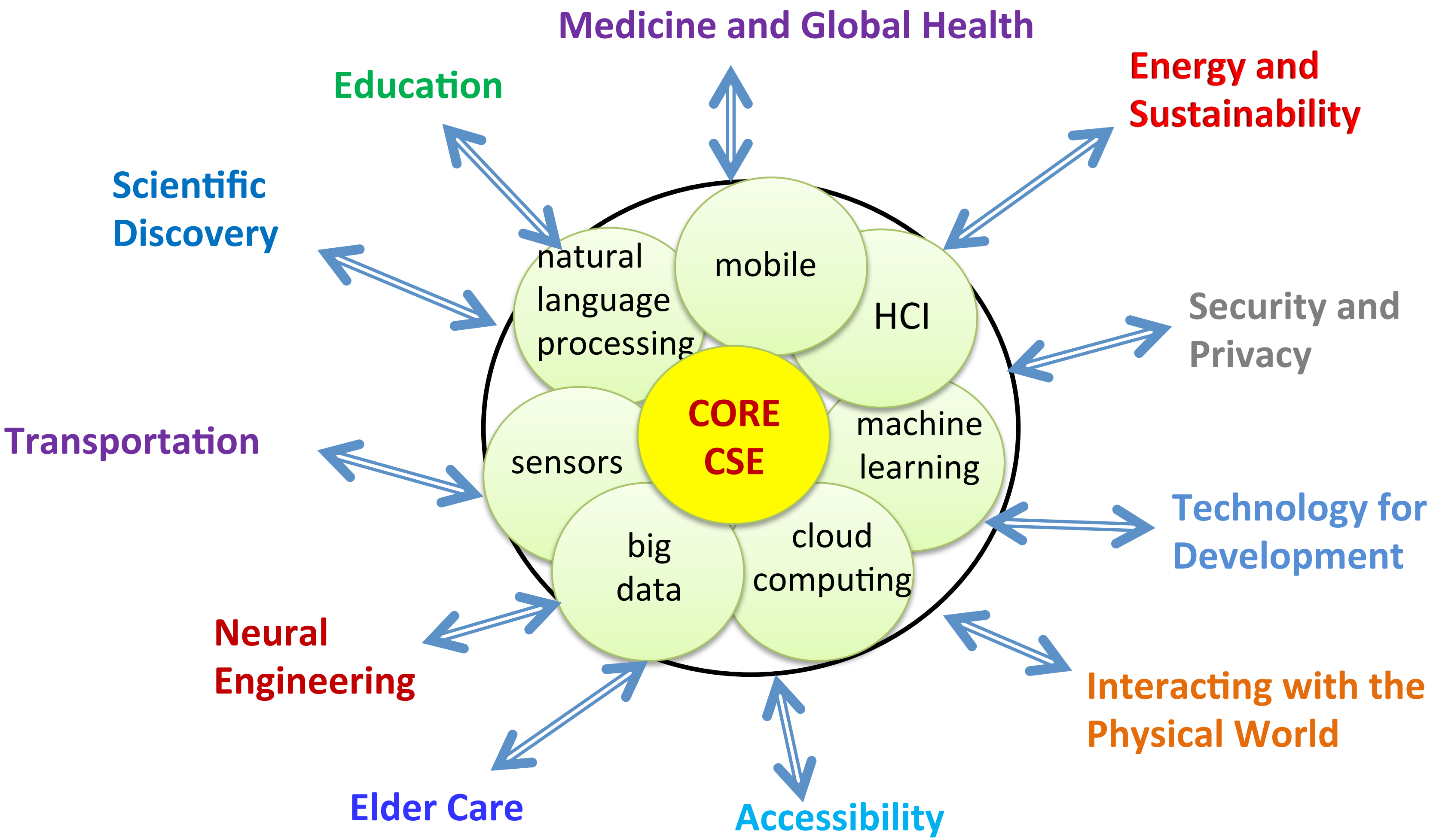
- NSF IGERT (interdisciplinary graduate education) award
- A signature 5-year center-scale award from two unnamed foundations to three unnamed universities
  - Stay tuned!





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

It drives an expansive view of the role of computing and of computer science



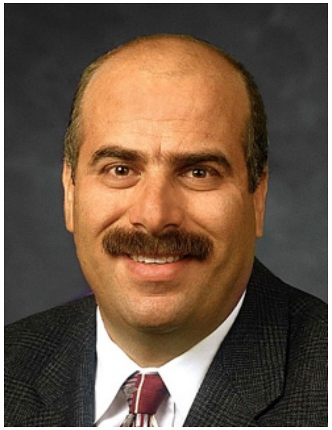
It represents the highest-impact leadership opportunity for NIAC

NORTHWEST INSTITUTE *for* ADVANCED COMPUTING

Pacific Northwest NATIONAL LABORATORY

eScience Institute

This was Moe's vision for NIAC



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

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