“It’s the Data, Stupid!”

Ed Lazowska

Bill & Melinda Gates Chair in
Computer Science & Engineering
University of Washington

Director, University of Washington eScience
Institute

127th MLA Annual Convention

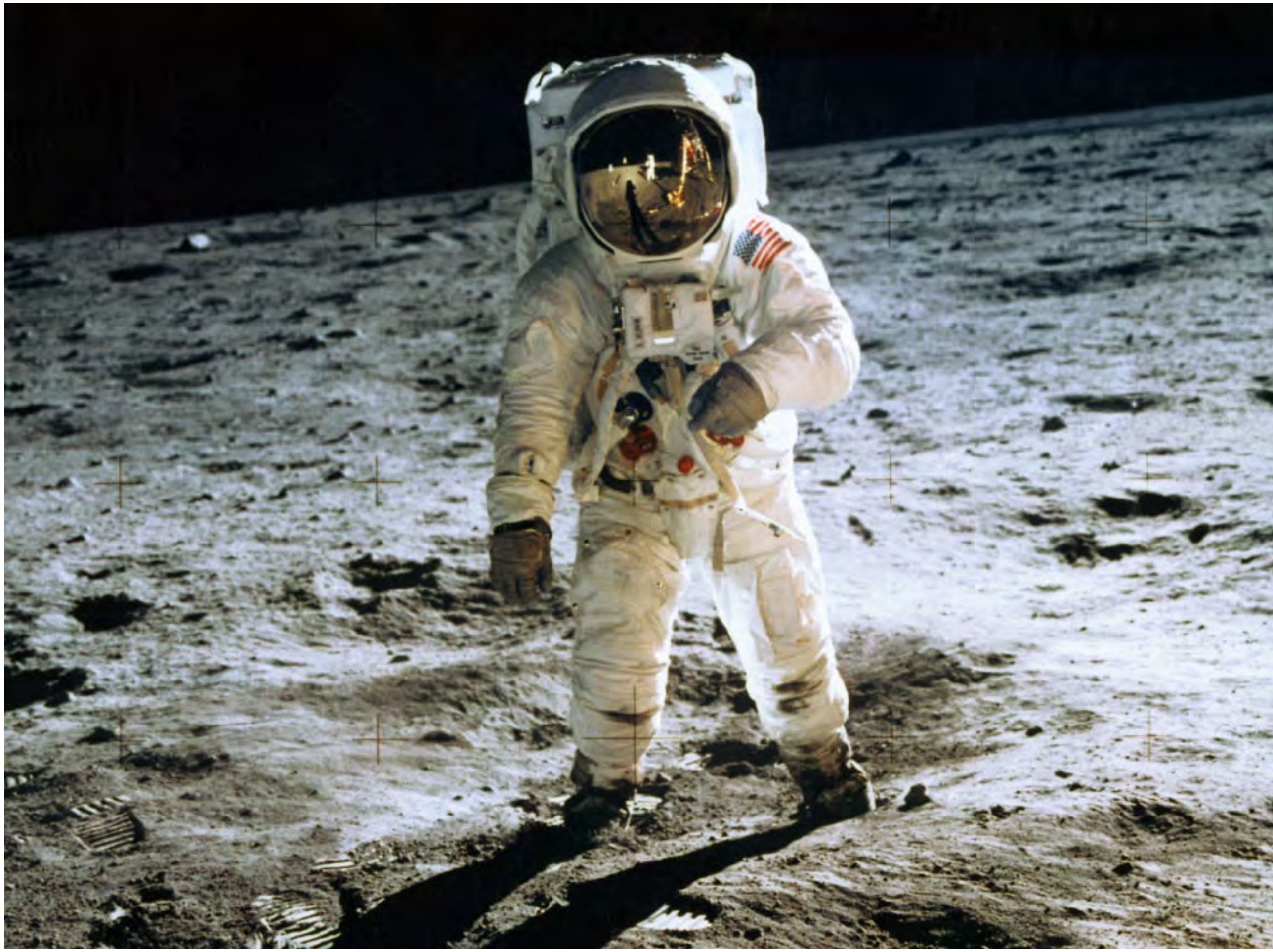
January 2012

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

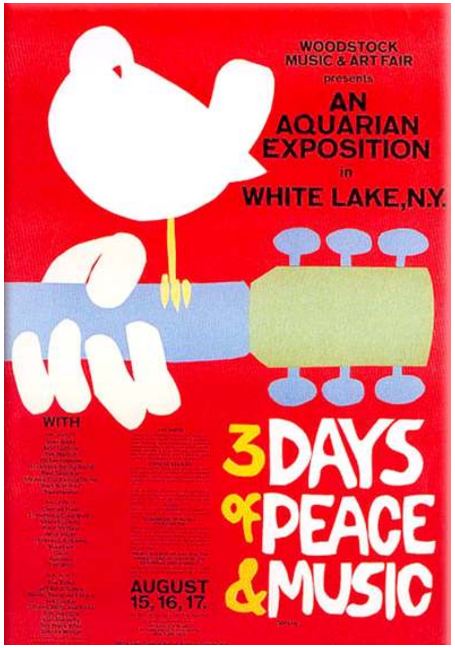
Forty years ago …

Credit: Peter Lee, Microsoft Research

USA
UNITED STATES

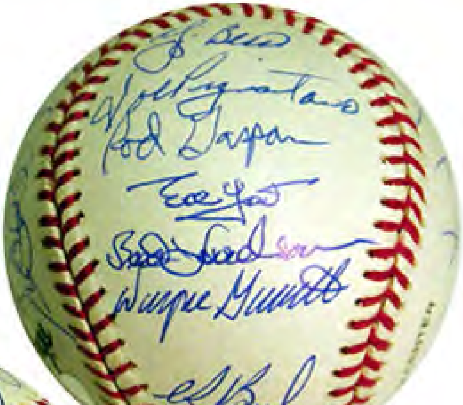


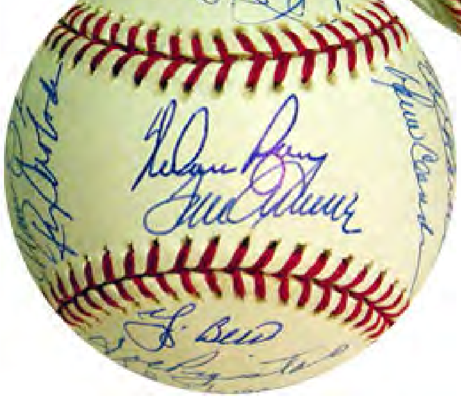
1969


WOODSTOCK MUSIC & ART FAIR
presents
AN AQUARIAN EXPOSITION
in
WHITE LAKE, N.Y.
WITH
AUGUST 15, 16, 17.
3 DAYS of PEACE & MUSIC

1969

1969 WORLD SERIES
NEW YORK METS

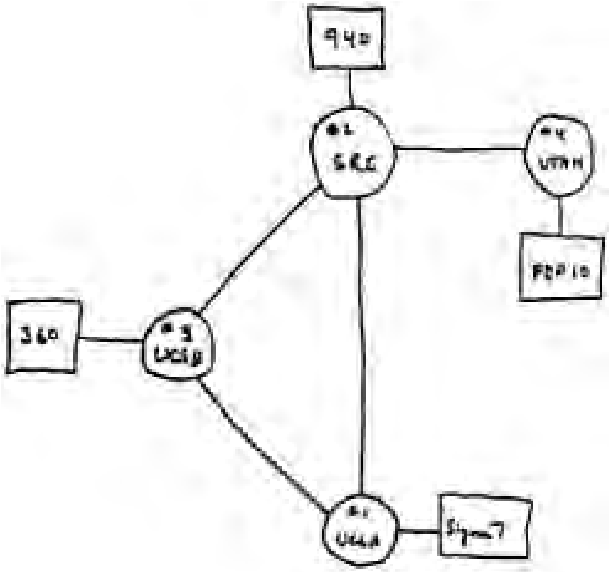

Rod Gaspar
Wayne Garrett


Nolan Ryan
Tom Seaver

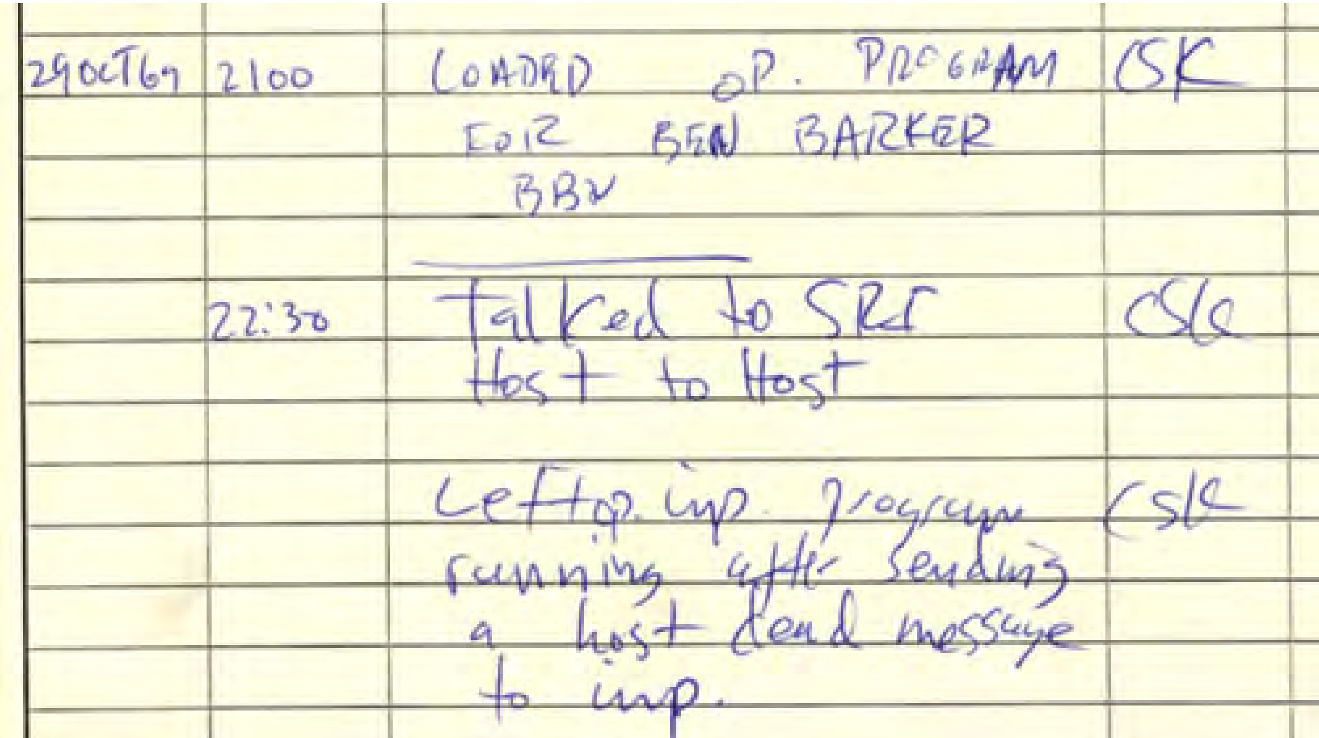
WORLD CHAMPIONS
NEW YORK METS
1969
30
RYAN
P

1969

Mets



THE ARPA NETWORK
DEC 1969
4 NODES



29OCT69	2100	LOADED OP. PROGRAM FOR BEN BARKER BBN	CSK
	22:30	Talked to SRI Host to Host	CSK
		Left op. imp. program running after sending a host dead message to imp.	CSK

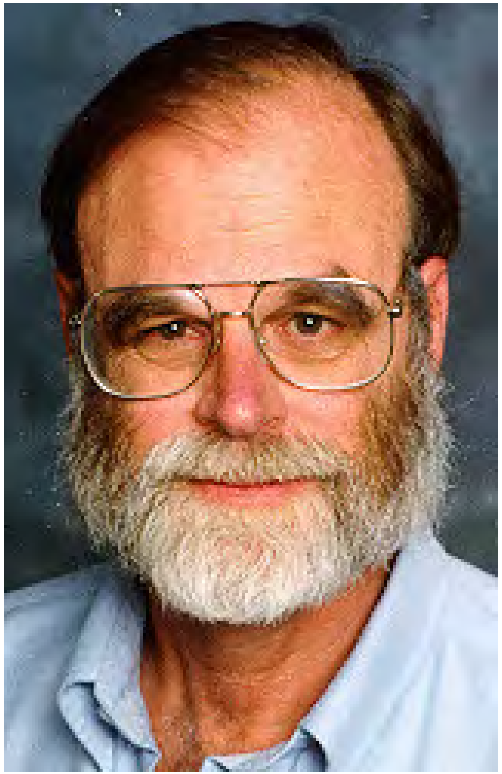
With forty years hindsight, which had the greatest impact?

- Unless you're big into Tang and Velcro (or sex and drugs), the answer is clear …

- And so is the reason …

EXPONENTIALS Я US

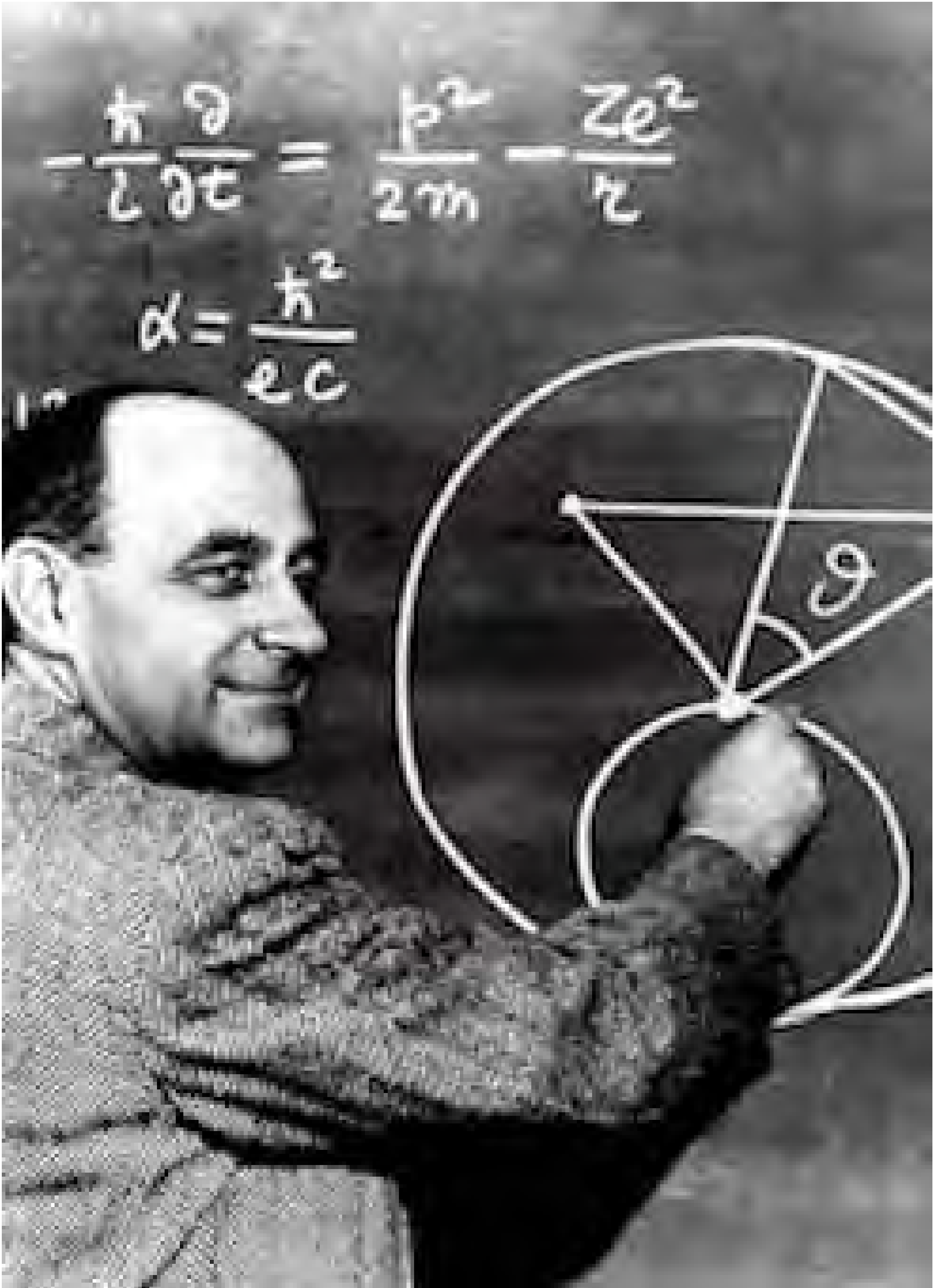
Today's exponential is data - eScience - data-intensive science and engineering



Jim Gray,
Microsoft Research

2020
SCIENCE
2020 COMPUTING SPECIAL: CHARTING THE INFORMATION EXPLOSION
REPRINTED FROM 23 MARCH 2006 VOL 440
nature
2020 VISION
How computers will change the face of science
Produced with support from
Microsoft Research
The FOURTH PARADIGM
DATA-INTENSIVE SCIENTIFIC DISCOVERY
EDITED BY TONY HEY, STEWART TANSLEY, AND KRISTIN TOLLE

Transforming science (again!)


$-\frac{\hbar}{i}\frac{\partial}{\partial t} = \frac{p^2}{2m} - \frac{Ze^2}{r}$
$\alpha = \frac{\hbar^2}{ec}$
ϑ

Theory

Experiment

Observation

Theory

Experiment

Observation

Theory

Experiment

Observation

Credit: John Delaney, University of Washington

Theory

Experiment

Observation

Computational Science

Theory

Experiment

Observation

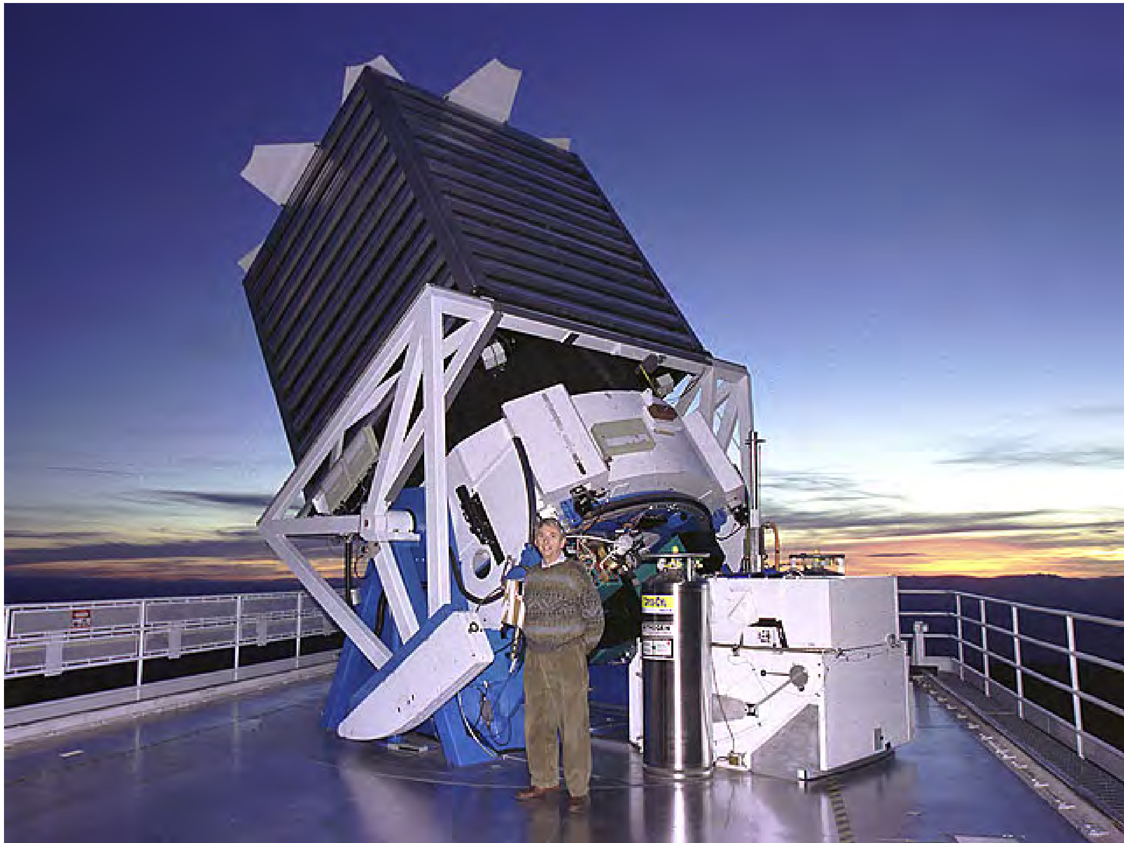
Computational Science

eScience

SLOAN DIGITAL SKY SURVEY

eScience is driven by *data* more than by cycles

- Massive volumes of data from sensors and networks of sensors (as well as from simulations)



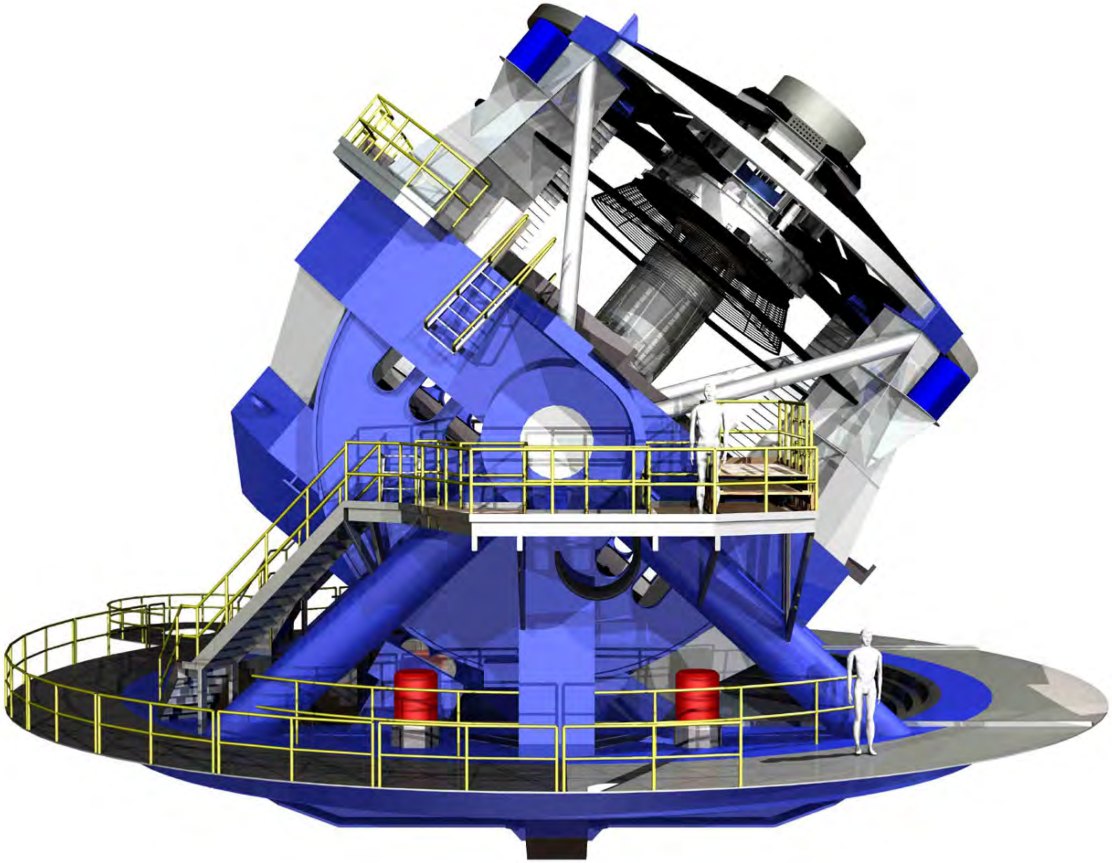
Apache Point telescope, SDSS

80TB of raw image data (80,000,000,000,000 bytes) over a 7 year period

Large Synoptic Survey Telescope (LSST)

40TB/day
(an SDSS every two days),
100+PB in its 10-year lifetime

400mbps sustained data rate between
Chile and NCSA



Large Hadron Collider

700MB of data per second, 60TB/day, 20PB/year

Illumina HiSeq 2000 Sequencer

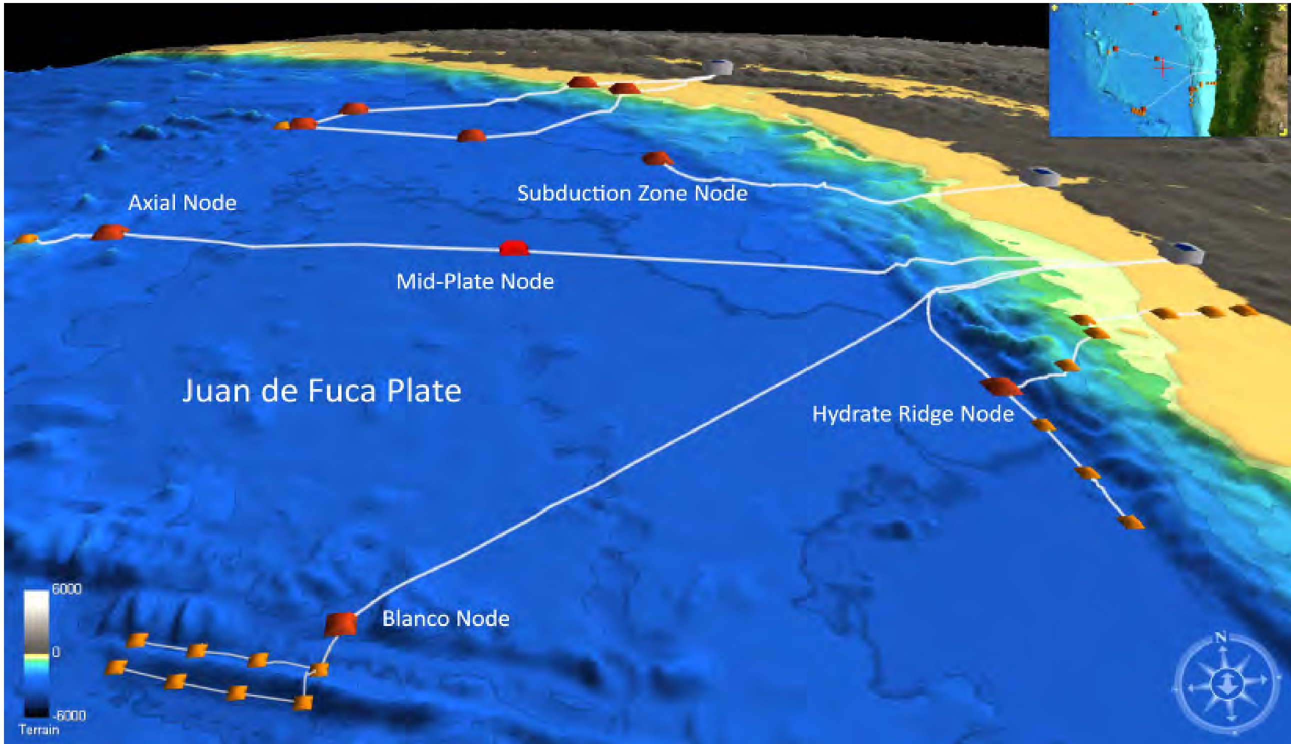
~1TB/day

Major labs have 25-100 of these machines

OCEAN OBSERVATORIES INITIATIVE

Regional Scale Nodes of the NSF Ocean Observatories Initiative

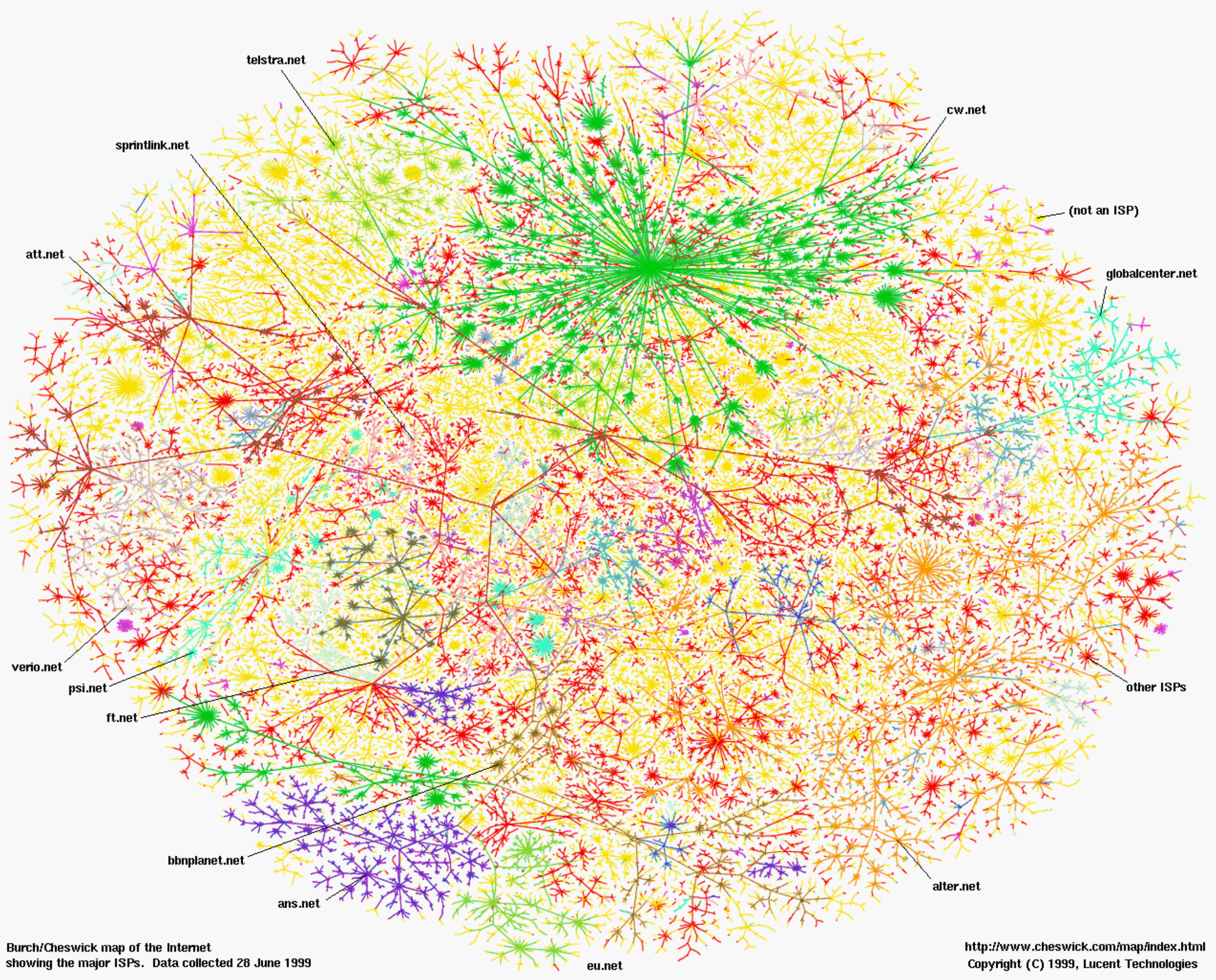
1000 km of fiber optic cable on the seafloor, connecting thousands of chemical, physical, and biological sensors



The Web

20+ billion web pages
x 20KB = 400+TB

One computer can
read 30-35 MB/sec
from disk => 4 months
just to read the web



Burch/Cheswick map of the Internet
showing the major ISPs. Data collected 28 June 1999

http://www.cheswick.com/map/index.html
Copyright (C) 1999, Lucent Technologies

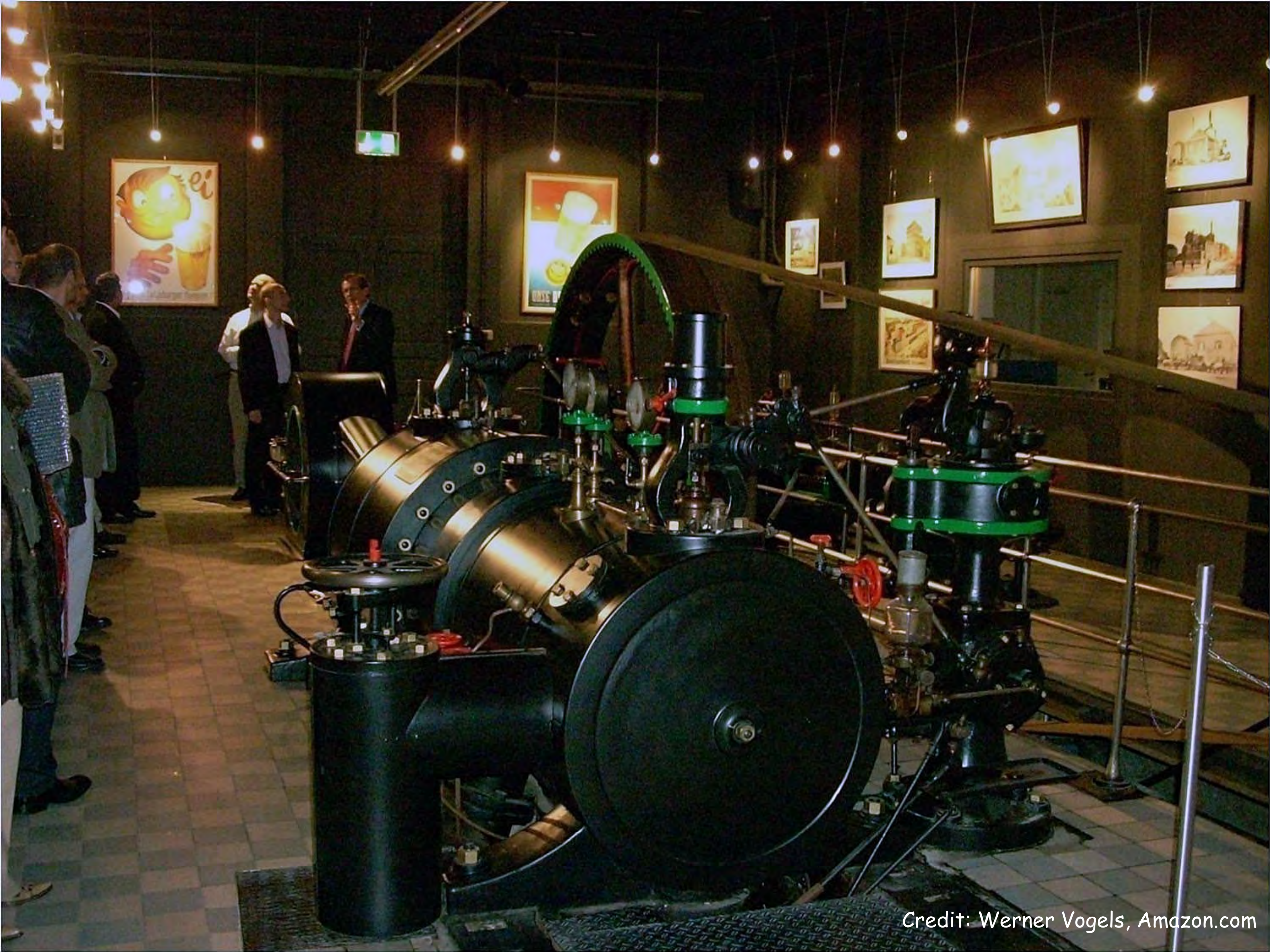
Point-of-sale terminals

eScience is about the *analysis* of data

- The automated or semi-automated extraction of knowledge from massive volumes of data
  - There's simply too much of it to look at
- It's not just a matter of volume
  - Volume
  - Rate
  - Complexity / dimensionality

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)



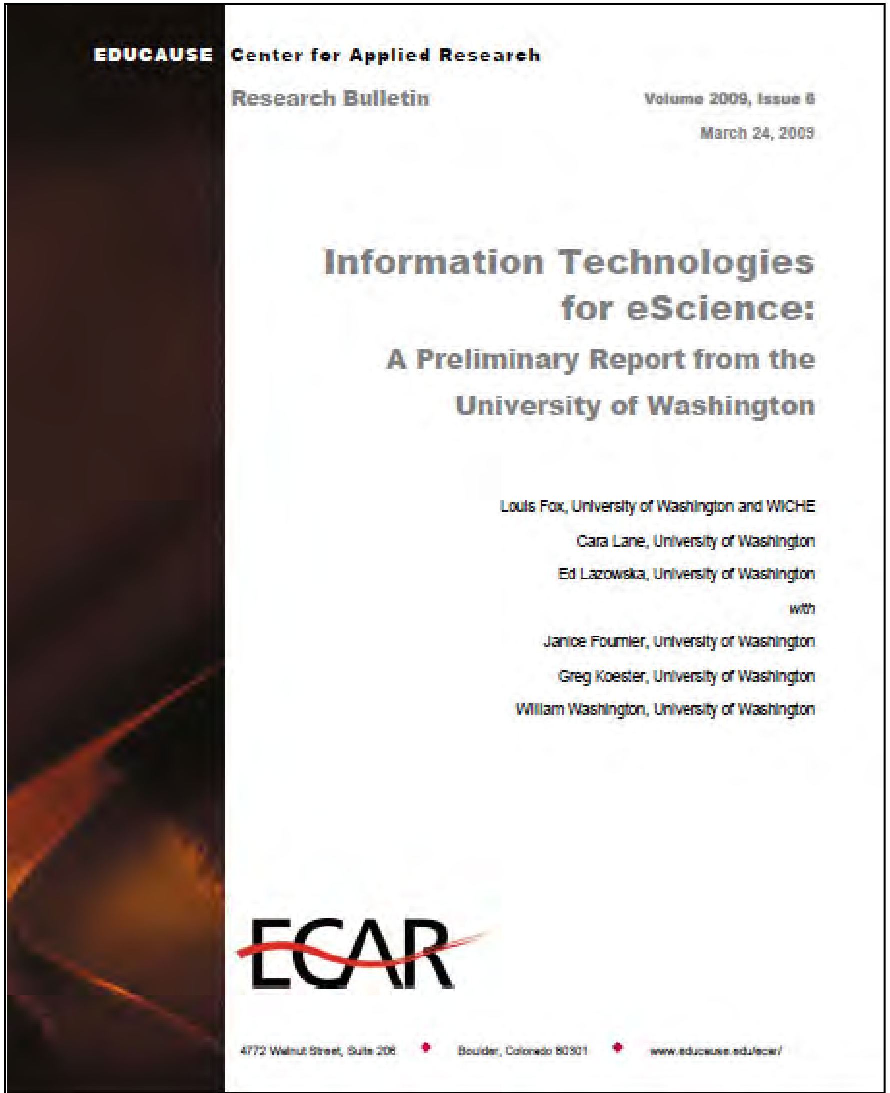
Credit: Werner Vogels, Amazon.com

eScience will be pervasive

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

eScience Institute

Top scientists across all fields grasp the implications of the looming data tsunami



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

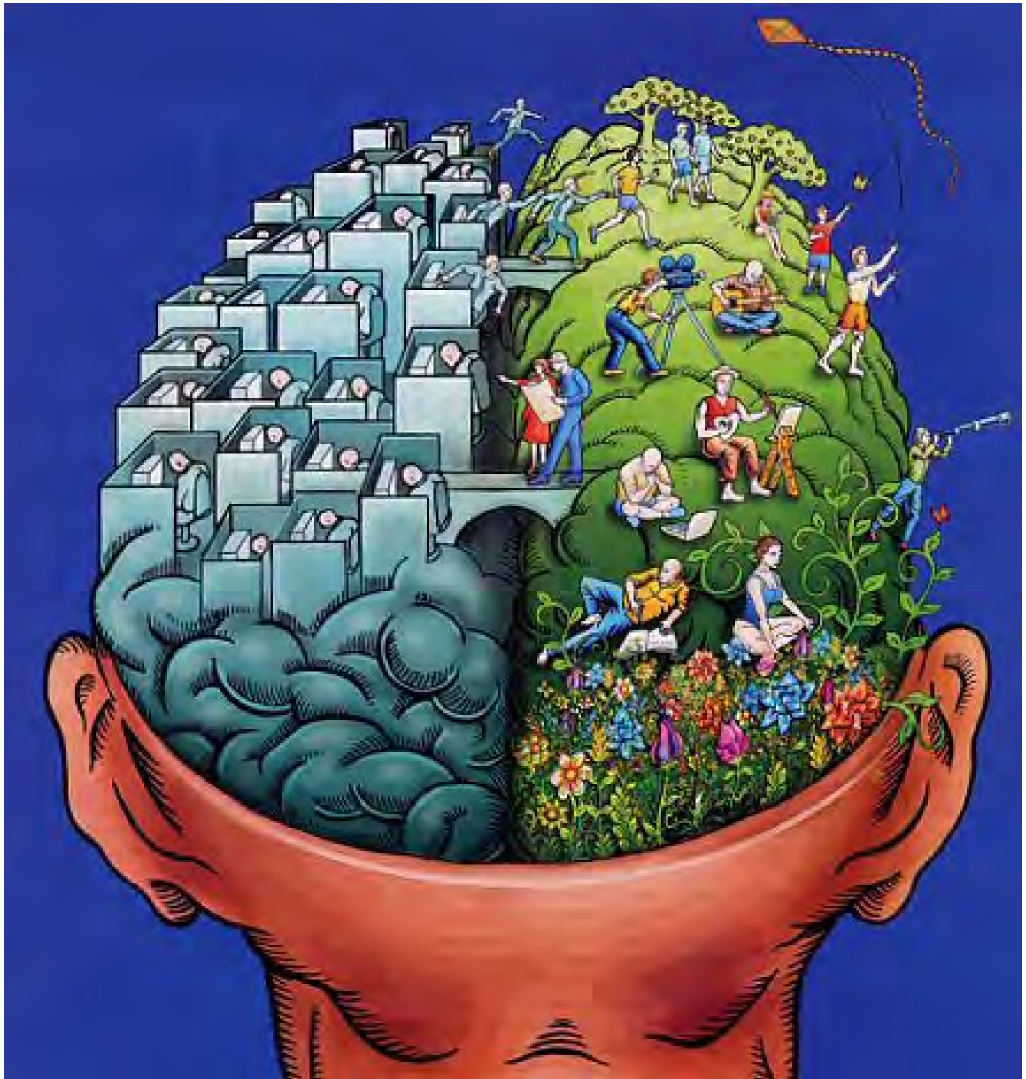
William Washington, University of Washington

ECAR

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

- Survey of 125 top investigators
  - "Data, data, data"
- Flat files and Excel are the most common data management tools
  - Great for Microsoft … lousy for science!
- Typical science workflow:
  - 2 years ago: 1/2 day/week
  - Now: 1 FTE
  - In 2 years: 10 FTE
- Need tools, tools, tools!

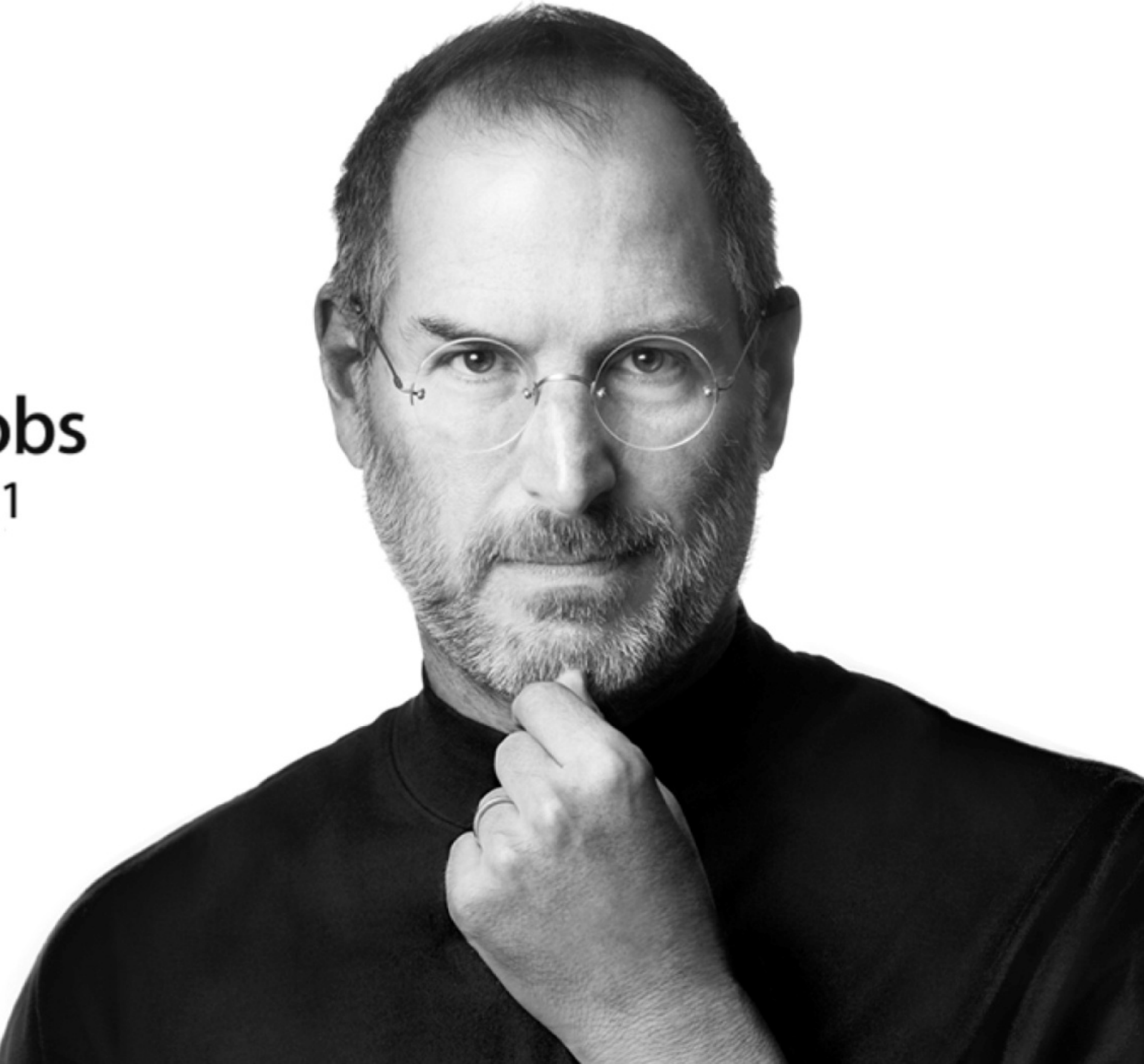
Side-note #1: The importance of using the whole brain

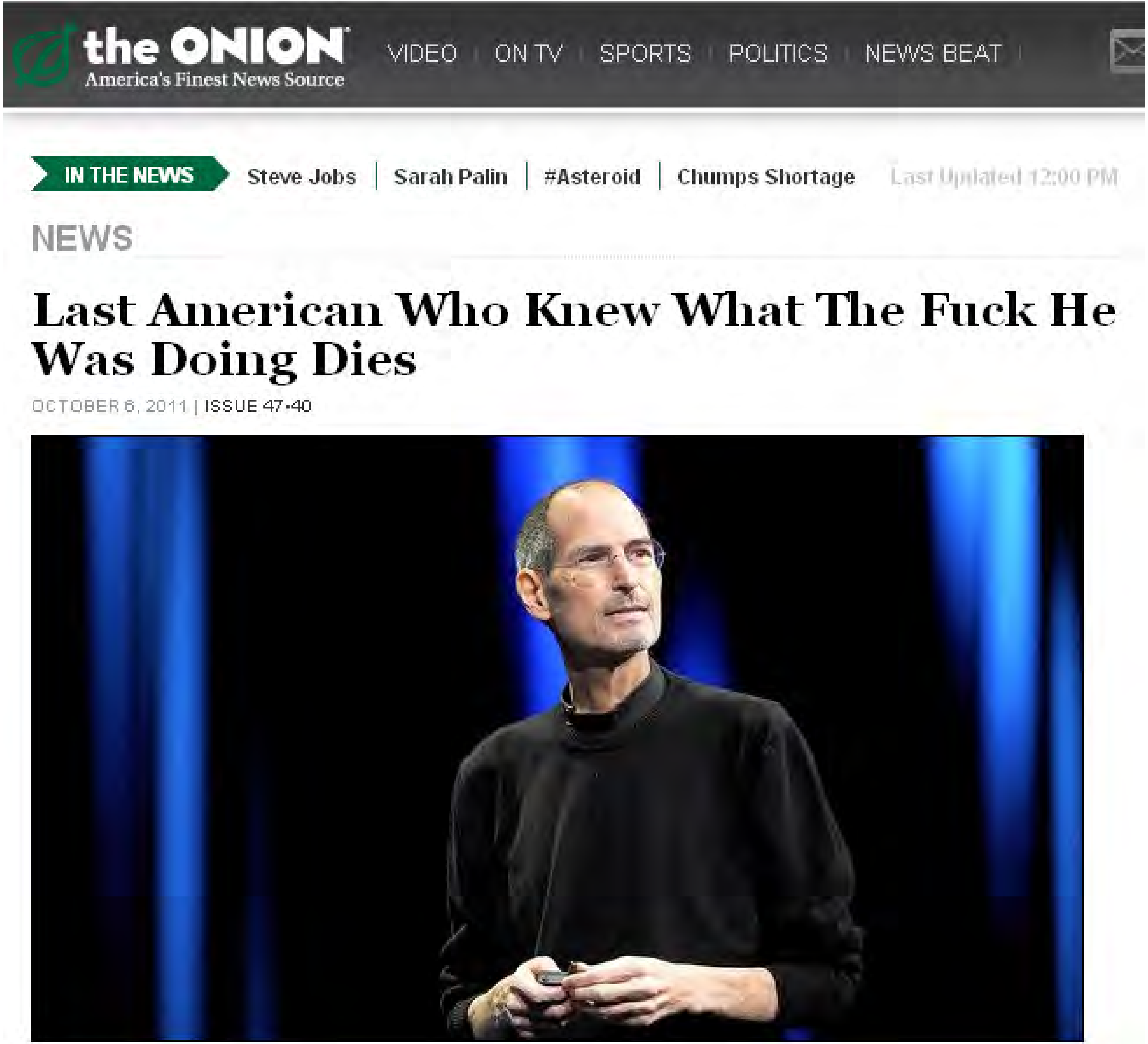


Credit: Julio Ottino, Northwestern

Steve Jobs

1955-2011

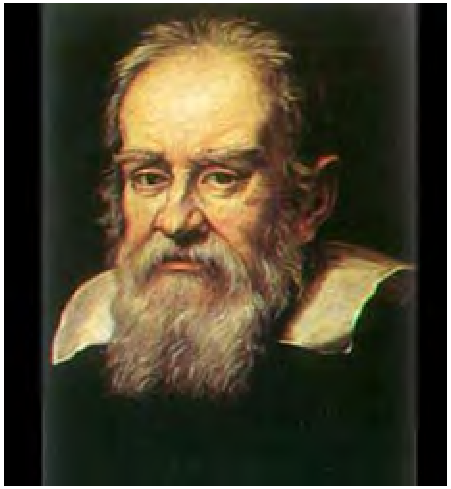




Last American Who Knew What The Fuck He Was Doing Dies

OCTOBER 6, 2011 | ISSUE 47•40

Harriot



Galileo

Credit: Julio Ottino, Northwestern

Side-note #2: The looming revolution of online learning

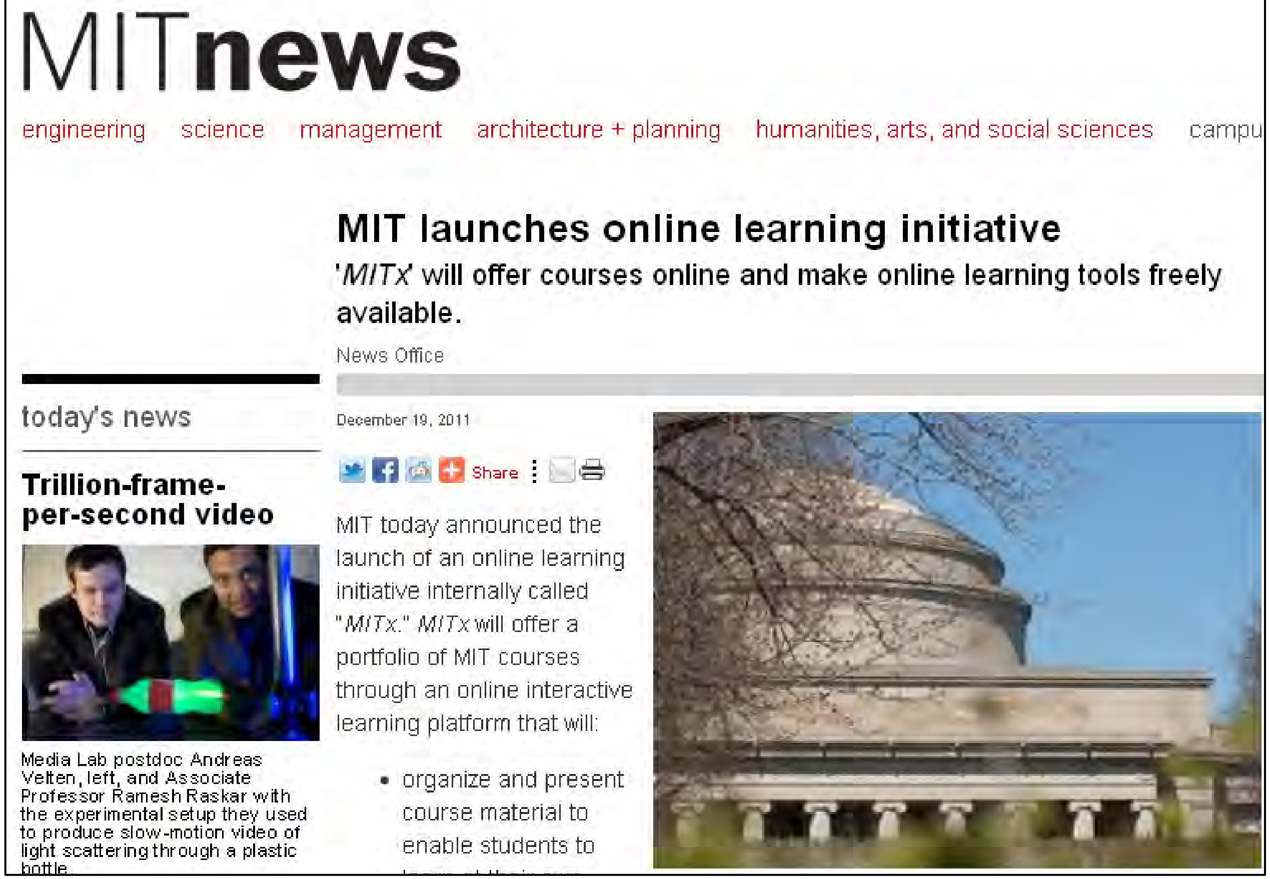
MIT
OCW

"MIT OCW gave me the opportunity to fulfill my dreams. Now is the time to return the favor. "
-Francisco, OCW Supporter, Spain

Please remember OCW in your end-of-year giving.
Your support makes a difference.

GIVE NOW >>

MITOPENCOURSEWARE
MASSACHUSETTS INSTITUTE OF TECHNOLOGY



MITnews

engineering science management architecture + planning humanities, arts, and social sciences campu

MIT launches online learning initiative

'*MITx*' will offer courses online and make online learning tools freely available.

News Office

December 19, 2011

Share

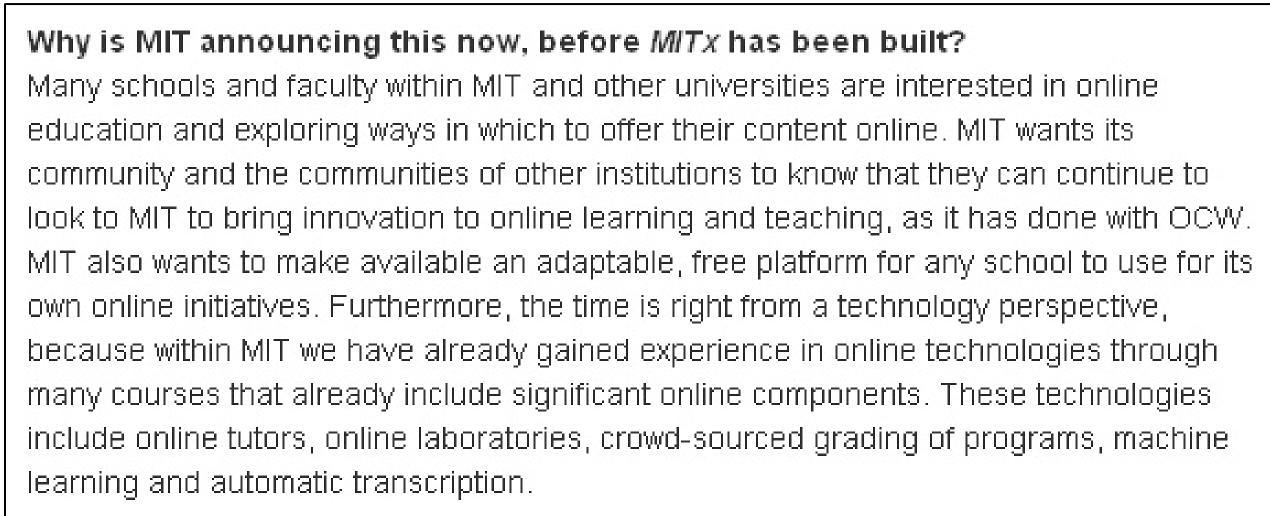
MIT today announced the launch of an online learning initiative internally called "*MITx*." *MITx* will offer a portfolio of MIT courses through an online interactive learning platform that will:

- organize and present course material to enable students to

today's news

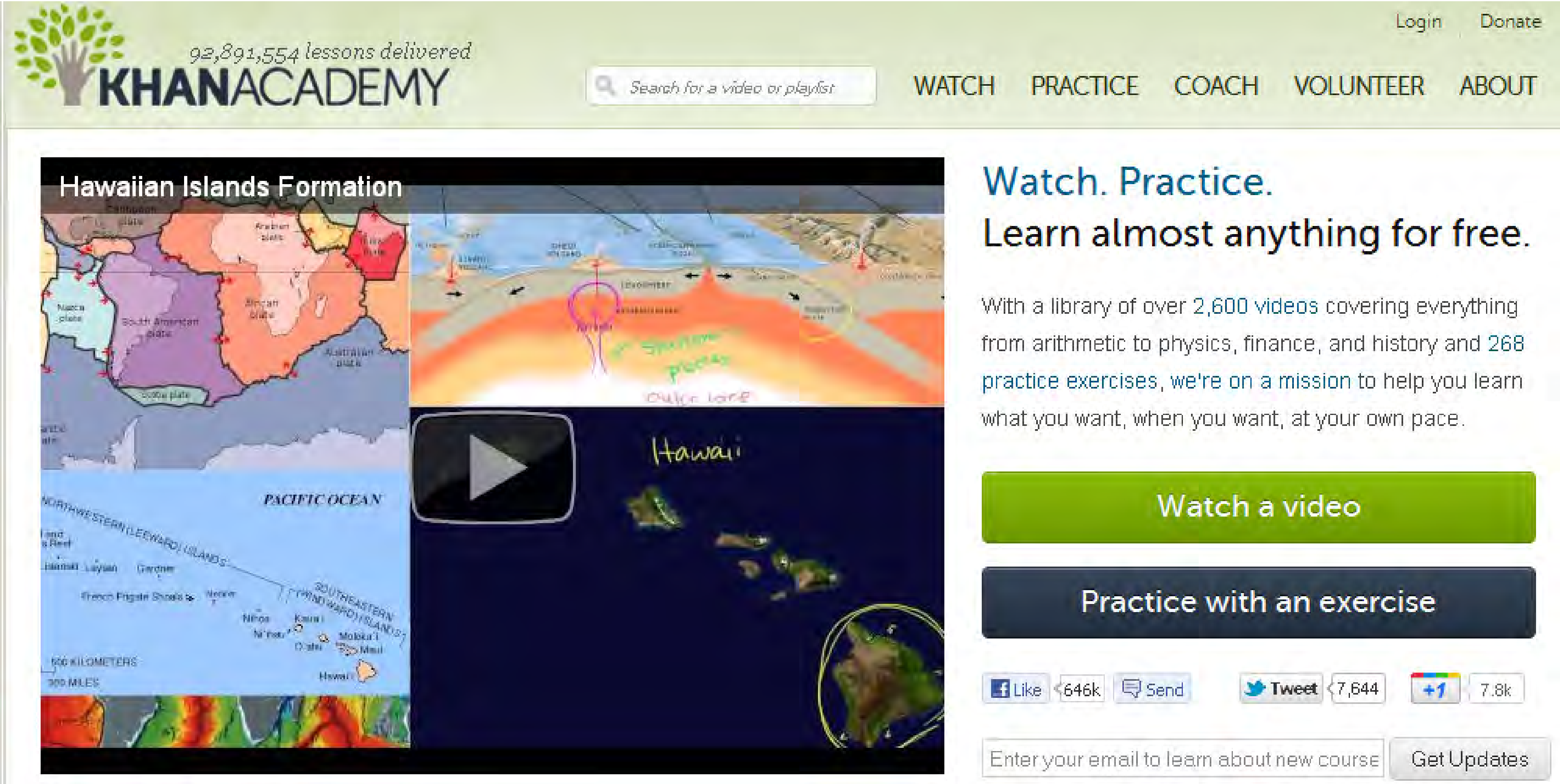
Trillion-frame-per-second video

Media Lab postdoc Andreas Velten, left, and Associate Professor Ramesh Raskar with the experimental setup they used to produce slow-motion video of light scattering through a plastic bottle.



Why is MIT announcing this now, before *MITx* has been built?

Many schools and faculty within MIT and other universities are interested in online education and exploring ways in which to offer their content online. MIT wants its community and the communities of other institutions to know that they can continue to look to MIT to bring innovation to online learning and teaching, as it has done with OCW. MIT also wants to make available an adaptable, free platform for any school to use for its own online initiatives. Furthermore, the time is right from a technology perspective, because within MIT we have already gained experience in online technologies through many courses that already include significant online components. These technologies include online tutors, online laboratories, crowd-sourced grading of programs, machine learning and automatic transcription.



92,891,554 lessons delivered

KHANACADEMY

Login Donate

Search for a video or playlist

WATCH PRACTICE COACH VOLUNTEER ABOUT

Hawaiian Islands Formation

Watch. Practice.

Learn almost anything for free.

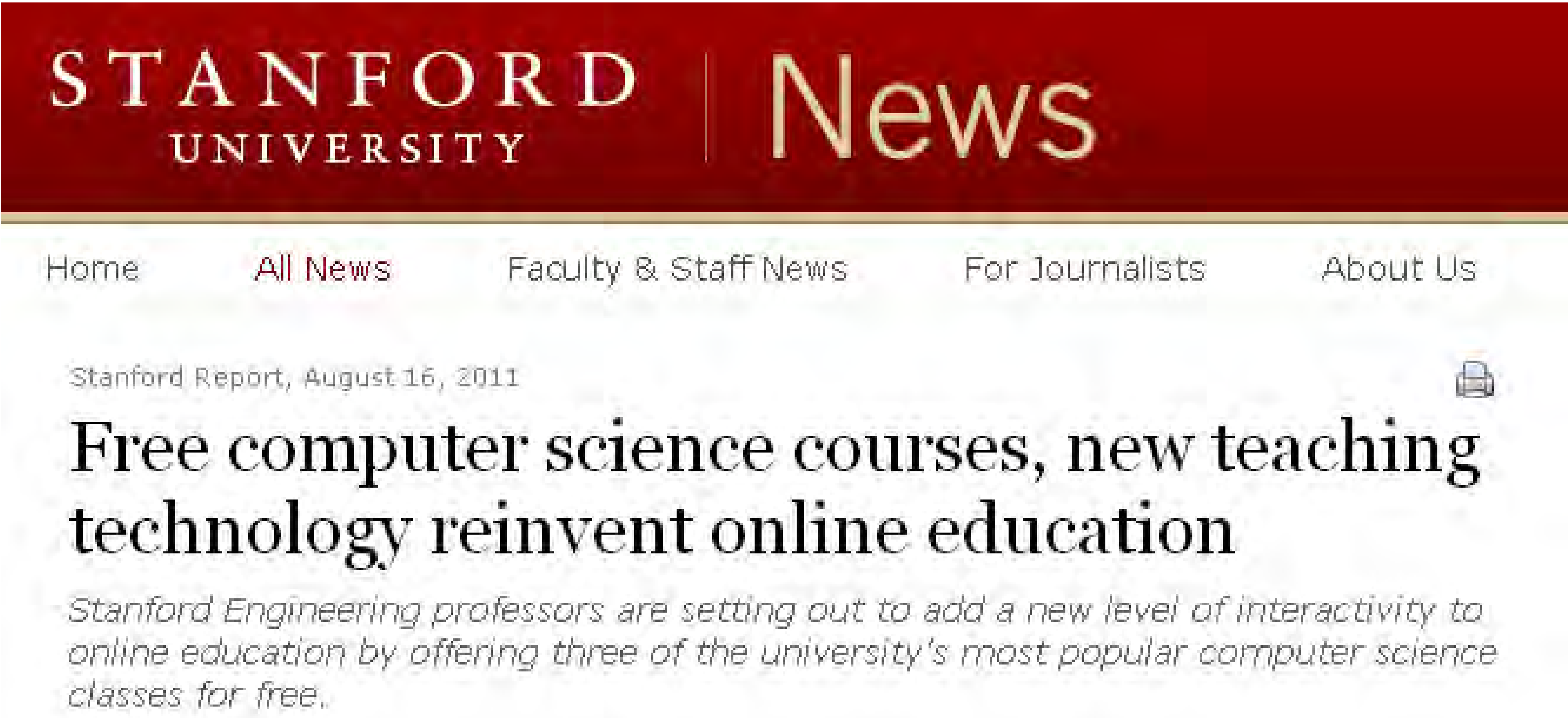
With a library of over 2,600 videos covering everything from arithmetic to physics, finance, and history and 268 practice exercises, we're on a mission to help you learn what you want, when you want, at your own pace.

Watch a video

Practice with an exercise

Enter your email to learn about new course

Get Updates



Stanford Report, August 16, 2011

Free computer science courses, new teaching technology reinvent online education

Stanford Engineering professors are setting out to add a new level of interactivity to online education by offering three of the university's most popular computer science classes for free.

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