



BIG DATA **THE MANAGEMENT REVOLUTION**



Erik Brynjolfsson
MIT Center for Digital Business
<http://digital.mit.edu/erik>

MIT Conference on Big Data, 12/12/12

Copyright © Erik Brynjolfsson



New Tools Beget Revolutions



Volume, Velocity, Variety

- Clickstream/Page views/Web transactions
- Web links/Blog references/Facebook
- Google/Bing/Yahoo Searches
- Email messages
- Mobile phone/GPS/Location data
- ERP/CRM/SCM transactions
- RFID (Radio Frequency Identification), Bar Code Scanner Data
- Real-time machinery diagnostics/engines/equipment
- Stock market transactions
- Twitter feeds
- Wikipedia updates
- Etc....

→ “Nanodata” and “Nowcasting”

3

Big Data

TERADATA

THE
PETABYTE
AGE

Copyright © 2012 Andrew McAfee

WIRED.CO.UK

HOME NEWS REVIEWS VIDEOS MAGAZINE

Latest Magazine | Magazine Archive | Read Magazine Features

Home > Magazine > 2012 > 08 > Features > The exabyte revolution

MAGAZINE

The exabyte revolution

By Neal Pollack | 16 July 12

CISCO Products & Services

Visual Networking Index (VNI)

The Zettabyte Era

4

Big Data

How Many? A Dictionary of Units of Measurement

© Russ Rowlett and the University of North Carolina at Chapel Hill

Metric Prefixes

To help the SI units apply to a wide range of phenomena, the 19th General Conference on Weights and Measures in 1991 extended the prefix yocto- at 10^{-24} (one septillionth). Here are the metric prefixes, with their numerical equivalents stated in the American system for

yotta- (Y-)	10^{24}	1 septillion
zetta- (Z-)	10^{21}	1 sextillion
exa- (E-)	10^{18}	1 quintillion
peta- (P-)	10^{15}	1 quadrillion
tera- (T-)	10^{12}	1 trillion

Copyright © 2012 Andrew McAfee

5

A Culture Clash

“I think we, as an industry, do a lot of talking... We expect to have open dialogue. It’s a culture of lunches.

Amazon doesn’t play in that culture. [It has] an incredible discipline of answering questions by looking at the math, looking at the numbers, looking at the data. . . .

That’s a pretty big culture clash with the word-and-persuasion-driven lunch culture, the author-oriented culture.”

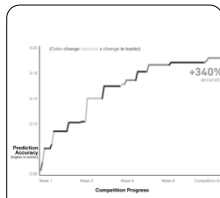
-- Madeline McIntosh, Random House’s President of Sales & Operations

Quoted In "Publish or Perish," *New Yorker*, 4/26/10

|| Data vs. HiPPOs



**Wine
Chemistry**



**Insurance
Underwriting**



**Housing
Sales**

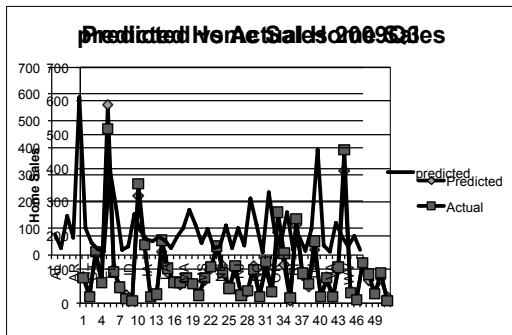
7

|| Data: Google Search Queries

- There are over 113 Billion search queries worldwide each month
 - 22% were in North America, 67% via Google. Increase of 41% annually.
 - Over 100 searches per searcher per month
 - Each query is a window into someone's intentions or interests.
 - Can we detect underlying consumer beliefs, perhaps "irrational exuberance" through online search data?
 - Shades of Montgomery Marvin and Hari Seldon
 - *Google Insights for Search* provides weekly or monthly reports on the query volume of any search phrase.
 - Related tool is Google Trends
- Can be drilled to country and state levels
 - Continuously improving it: Now available for some MSAs
- Collect query data since Jan 1, 2004

Copyright (c) Erik Brynjolfsson

8



In the September, 2009 draft of NBER paper (submitted for workshop), we used our model to estimate sales in Q3 2009.

We predicted the average home sales for the entire nation would **increase by 6.4%**

The National Association of Realtors predicted an **8.1% increase**.

On Nov. 24, 2009, the housing sales for Q3 were announced ...

... as increasing 6.1%.

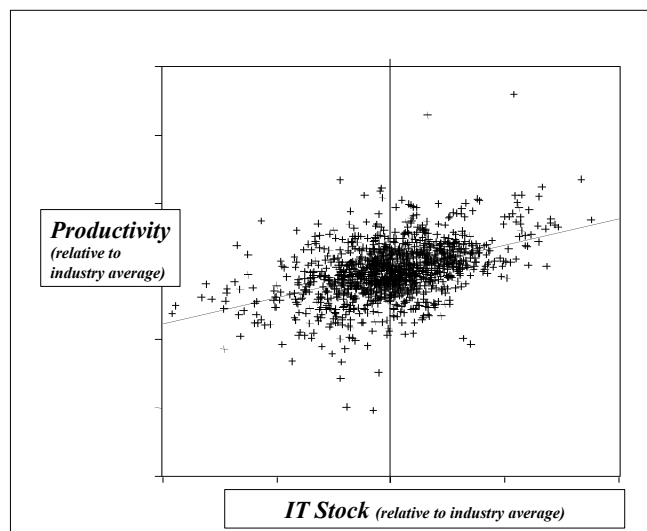
**Nanodata produced a 66% improvement over the baseline model*

Copyright (c) Erik Brynjolfsson

Source: Brynjolfsson and Wu, 2009

9

The Plural of Anecdote is Data

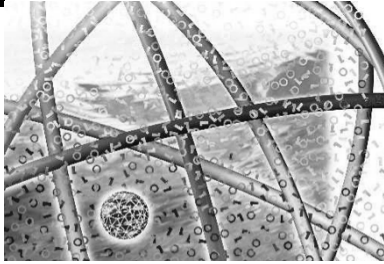


(Source: Brynjolfsson and Hitt)

Copyright (c) Erik Brynjolfsson

10

Data Driven Decision-makers are Winning



✓ Data-Driven Decision-makers:
4% higher productivity

✓ *Data-Driven Decision-makers:*
6% greater profitability and 50%
higher market value from IT



Source: Brynjolfsson, Hitt and Kim, 2011

Copyright (c) Erik Brynjolfsson

11

Data-Driven Decision-Making (DDD)

1. How are decisions made for the creation of a new product or service? (1 to 5 scale: Experience and expertise=1, Data=5)
2. To what extent do the following statement describe the work practices and environment of your entire company
 - 2a. We depend on data to support our decision making (1: Describes not at all, 5: Completely describes)
 - 2b. We have the data we need to make decisions (1: Describes not at all, 5: Completely describes)

SOURCE 2009 Digital Advantage survey by MIT CDB and McKinsey

12

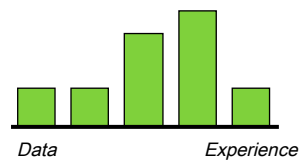
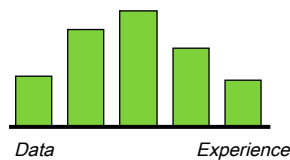
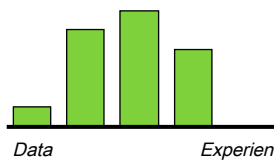
Data-driven decision making I: Typical basis for a new product/service

Percent of respondents

**Minerals, Oil & Gas,
Utilities, and Construction**

Manufacturing

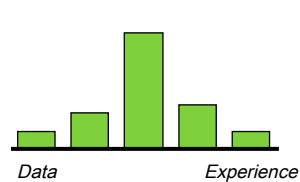
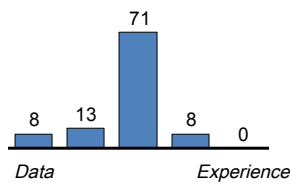
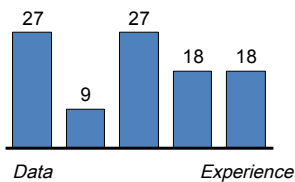
**Wholesale/Retail Trade,
Transport, Accommod./Food**



Information

Finance and Insurance

**Professional and
Other Services**



13

The Effect of DDD on Productivity

$$\begin{aligned} \ln(\text{Sales})_{it} = & \beta_0 + \beta_1 \ln(\text{Materials})_{it} + \beta_2 \ln(\text{Physical Capital})_{it} + \\ & \beta_3 \ln(\text{IT Labor})_{it} + \beta_4 \ln(\text{Non-IT Labor})_{it} + \beta_5 (\text{DDD})_i \\ & + \text{Other controls} \end{aligned} \quad > 0 ?$$

i: firm

t: year (2005-2009)

Sales, Physical Capital, Employees from Compustat

IT Labor from a job-posting site (Tambe and Hitt, 2008)

Non-IT Labor = Employees – IT Labor

Other controls = 1.5 digit NAICS industry, year,

employees' human capital

(importance of typical employee's education, % of employees using PC/Emails, and/or Avg. workers' wage)

Copyright (c) Erik Brynjolfsson

14

Productivity and DDD: OLS and IV

	OLS	IV
DDD	0.0475** (0.019)	0.064* (0.035)
Ln (Material)	0.501 *** (0.042)	0.504*** (0.034)
Ln(Physical Capital)	0.0991*** (0.023)	0.0979*** (0.023)
Ln (Non-IT Employee)	0.224 *** (0.032)	0.224*** (0.032)
Ln (IT-Employee)	0.0852*** (0.022)	0.0844*** (0.022)
Industry and Year Control	Yes	Yes
R-squared	0.92	0.92
Overid Test: Hansen' s J		0.68
Hausman Test		0.58

Robust standard errors were clustered around firms.

****p<0.01, **p<0.05, *p<0.1. Industry classification was based on NAICS 2 digit for manufacturing and 1 digit for other industries.*

15

DDD also improved the other performance measures

1. Return on Assets
 - Pretax Income per total assets
2. Return on Equity
 - Pretax Income per equity
3. Asset Utilization
 - Output per total assets
4. Market Value
 - Market to Book Ratio

Copyright (c) Erik Brynjolfsson

16

Big Data:

The Catalyst for Four Drivers of Innovation

Measure

Experiment

Share

Replicate

17

Synergies Amplify The Benefits

Replicate

Measure

Experiment

Share

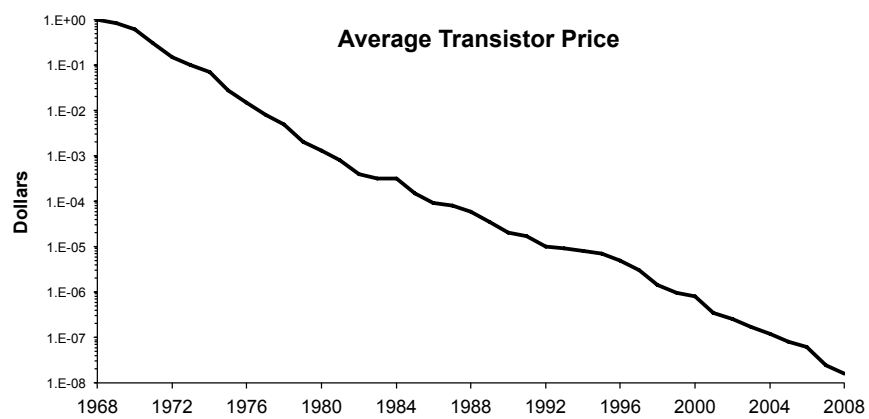
A new kind of R&D

18

|| The Future of Prediction

19

|| The Falling Costs of Measurement, Experimentation, Sharing and Replication

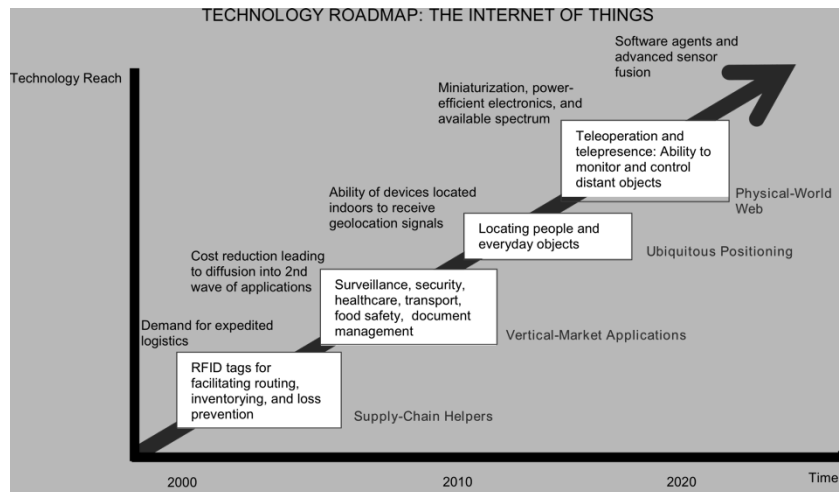


Copyright (c) Erik Brynjolfsson

20

The Internet of Things

By 2015 there will be 1 trillion sensors linking the physical and digital worlds merging to become an "Internet of Things"



Source: SRI Consulting Business Intelligence
Copyright (c) Erik Brynjolfsson

21

To learn more about related research,
please visit:

<http://digital.mit.edu>

Big Data: <http://digital.mit.edu/bigdata>

Erik's papers: <http://digital.mit.edu/erik>