The Future of Data Management

PK Agarwal
Dean and CEO
pk.agarwal@neu.edu

Northeastern University
Silicon Valley

Agenda

1. Four Stages of Data Management
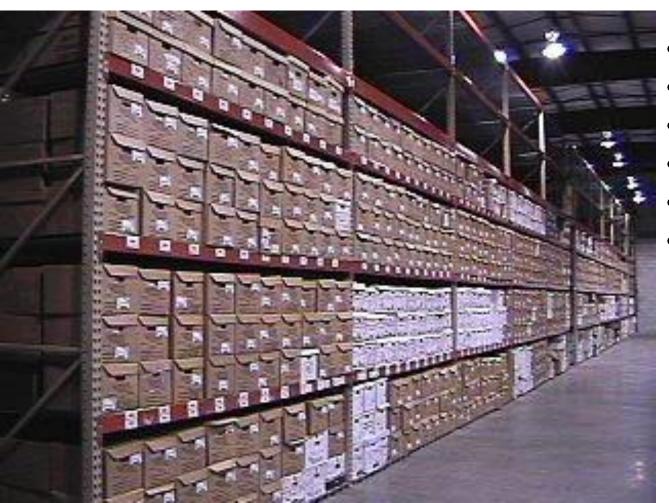
2. A Professional Road Map

3. Q/A

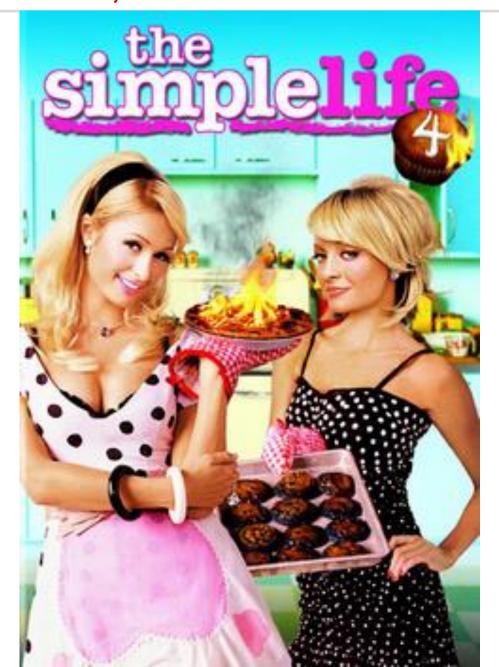# World Fair 1893 - Chicago

The Future – Eyes of 1893

- Taxation reduced to a minimum
- Entire world open to trade
- No need for standing armies
- People will live to be 150
- Government would have become more simple because true greatness always tends toward simplicity
- Free mail delivery, no mortgages, removal of divorce laws…..

Phase I – We Thought it was Big Data!

- Safe and Secure
- Clear Rules
- Governance
- Scale
- Efficiency
- 1,000+ Years in the making

Phase II – Born Digital

"As we move into the electronic era of digital objects it is important to know that there are new barbarians at the gate and that we are moving into an era where much of what we know today, much of what is coded and written electronically, will be lost forever. We are, to my mind, living in the midst of digital Dark Ages. "

Terry Kuny, National Library of Canada

Phase II – Born Digital

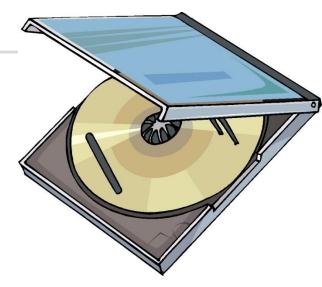Would the following images be available to the generation 100 years from now?

These are Available!!

EARTHQUAKE WRECKED BUILDINGS ON HOWARD ST.

RELIEF LINE AT OAKLAND.

CLINTON JOHNSON, PUB.
LOS ANGELES, CAL.

The year is 2045, and my grandchildren (as yet unborn) are exploring the attic of my house (as yet unbought). They find a letter dated 1995 and a CD-ROM (compact disk). The letter claims that the disk contains a document that provides the key to obtaining my fortune (as yet unearned). My grandchildren are understandably excited, but they have never seen a CD before—except in old movies—and even if they can somehow find a suitable disk drive, how will they run the software necessary to interpret the information on the disk? How can they read my obsolete digital document?
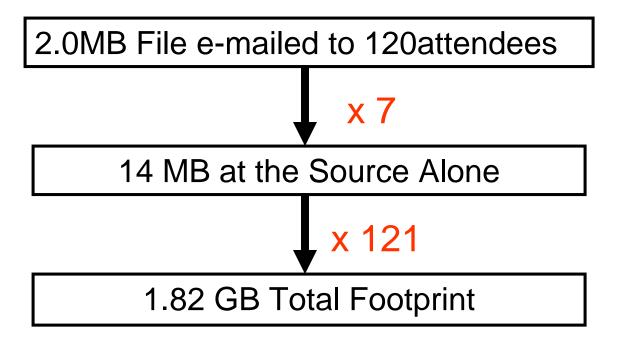
Jeff Rothenberg, "Ensuring the Longevity of Digital Documents"

Digital Dark Age

- Print represents < 0.01% of new information being created

- How do we preserve information and make it available to future generations

> Thanks to Digitization, we are Losing Our Institutional and Personal Memories

Data Redundancy

2.0MB File e-mailed to 120attendees

x 7

↓

14 MB at the Source Alone

x 121

↓

1.82 GB Total Footprint

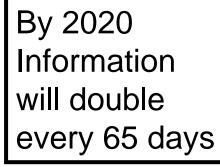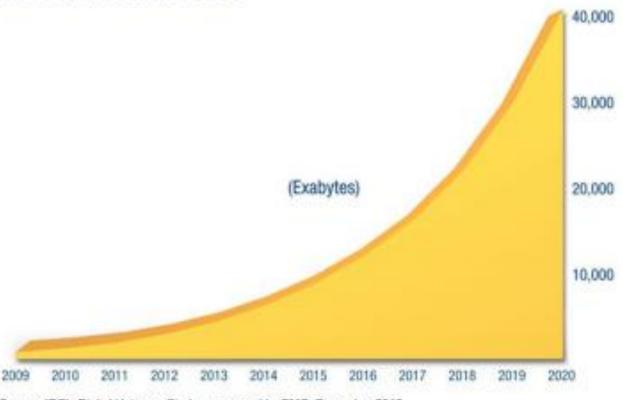# Dramatically Reduced Cost of Storage

Phase II – Born Digital

- eDiscovery 2006 reinforced Governance of electronic data (same as records?)
- Records Management + Information Governance

Expanding Data Universe

The Digital Universe: 50-fold Growth from the Beginning of 2010 to the End of 2020

By 2020 Information will double every 65 days

(Exabytes)

40,000

30,000

20,000

10,000

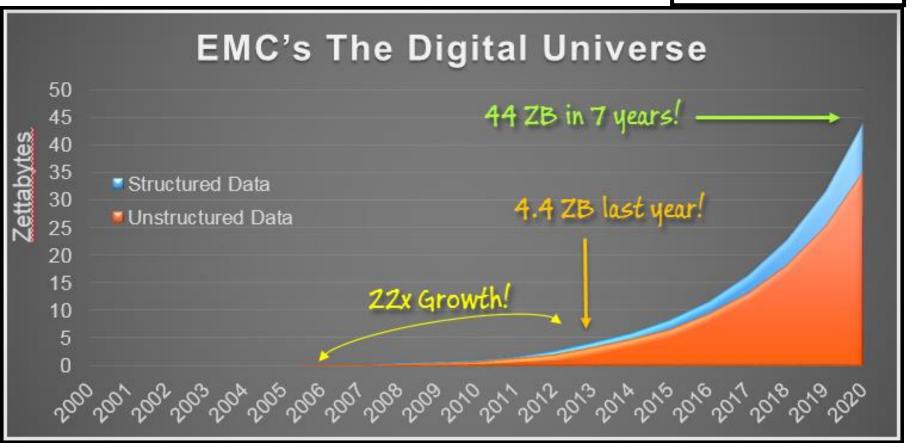2009 2010 2011 2012 2013 2014 2015 2016 2017 2018 2019 2020

Source: IDC's Digital Universe Study, sponsored by EMC, December 2012

Expanding Data Universe

By 2020 Information will double every 65 days

EMC's The Digital Universe

44 ZB in 7 years!

4.4 ZB last year!

22x Growth!

- Structured Data
- Unstructured Data

Zettabytes

Phase III – It's a Mad World

- **Majority of Data is Created Outside of the Organization**
- Massive Growth in Unstructured Data
  - eMail, Graphics, Pictures, Audio, Video, VoIP
- Further Growth of Structured Data
  - Sensors, RFID
  - More Transactions
    - Once a month vs. EBT Card
- Everyone is a Content Creator
  - Blogs, Flickr, Social Networks, YouTube

Big Data

- Analytics
- 360 view of the customer
- Intelligence
- Predictive

Franchise Tax Board Field Office

★★★★★ 10 reviews 📊 Details

Write a Review

Public Services & Government ✏ Edit

300 S Spring St ✏ Edit
Ste 5704
Los Angeles, CA 90013
Downtown

Get Directions
📞 (800) 852-5711
🔗 ftb.ca.gov/aboutFTB/tscadd.sh...

Mural in State bldg, note fountain... by Jef H.

Yelp!!!!

Exploding Data Universe

- 1,000,000,000,000,000,000,000 individual web pages

- Social Media Sample
  - Facebook - 500,000 comments/minute
  - YouTube – 30 hours/minute
  - Twitter, LinkedIn, Blog posts, Searches

- IoT is coming!

Big Data Expectations
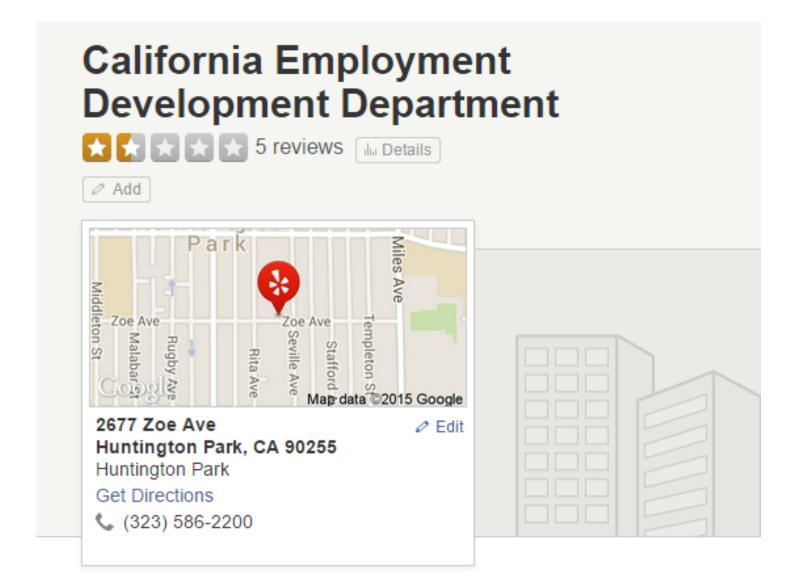
Data Analytic Correlations (R)

- 1. Kenny Chesney
- 2. George Strait
- 3. Reba McEntire
- 4. Tim McGraw
- 5. Jason Aldean
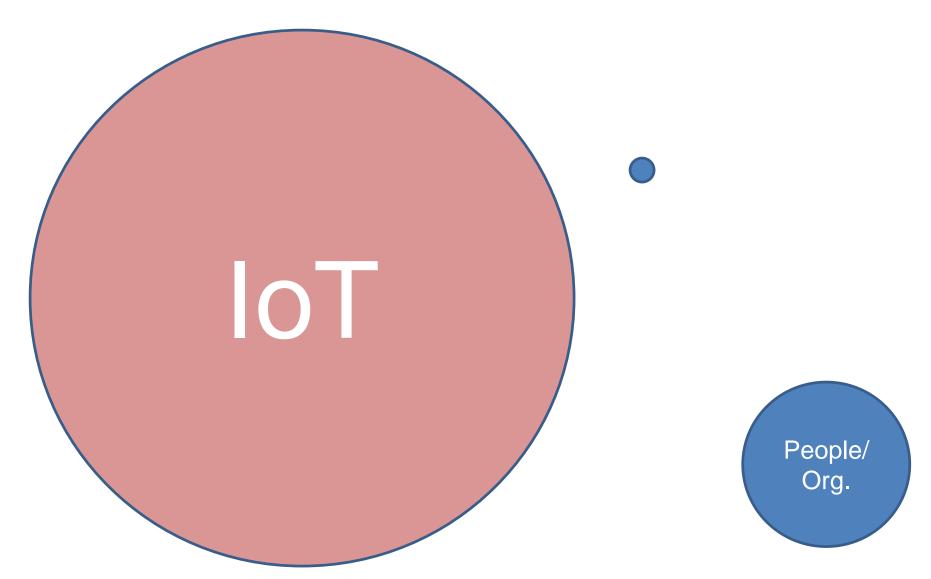
- 6. Blake Shelton
- 7. Shania Twain
- 8. Kelly Clarkson
- 9. Pink Floyd[7]
- 10. Elvis Presley

Data Analytic Correlations (D)

- 1. Rihanna

- 2. Jay-Z

- 3. Madonna

- 4. Lady Gaga

- 5. Katy Perry

- 6. Snoop Dogg

- 7. Chris Brown

- 8. Usher

- 9. Eminem

- 10. Bob Marley

http://notes.variogr.am/post/26869688460/how-well-does-music-predict-your-politics

Phase IV - Infinite Data

Some 30 billion objects may be connected to the
Internet of Things[1] by 2020.

**26 billion–30 billion
objects in 2020**

**7 billion–10 billion
objects in 2013**

*~15–20%
growth
annually*

[1]A networking of physical objects via embedded devices that collect and/or transmit information.

Source: Forecasts derived from ABI Research; expert interviews; Gartner; IDC; McKinsey analysis

Consumer vs. Industrial

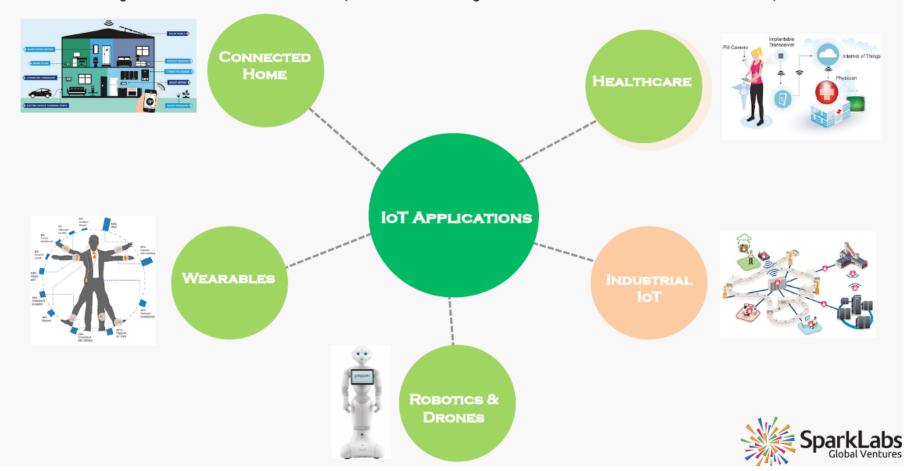We see the landscape generally divided between consumer and industrial applications whether hardware, software or both. The consumer innovations are primarily hardware products and industrial have been primarily software (i.e. artificial intelligence, data analysis). Most connected home, wearables and robotics/drones products have been consumer facing while healthcare has been split with a leaning towards with more hardware-focused products.
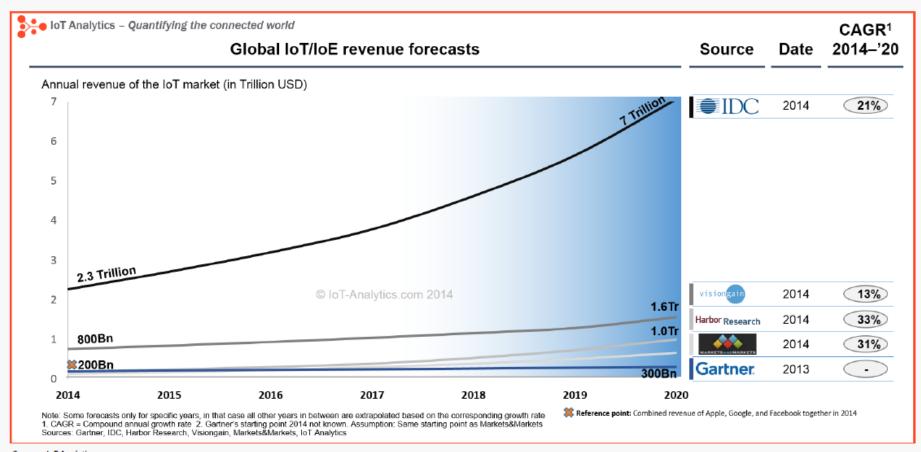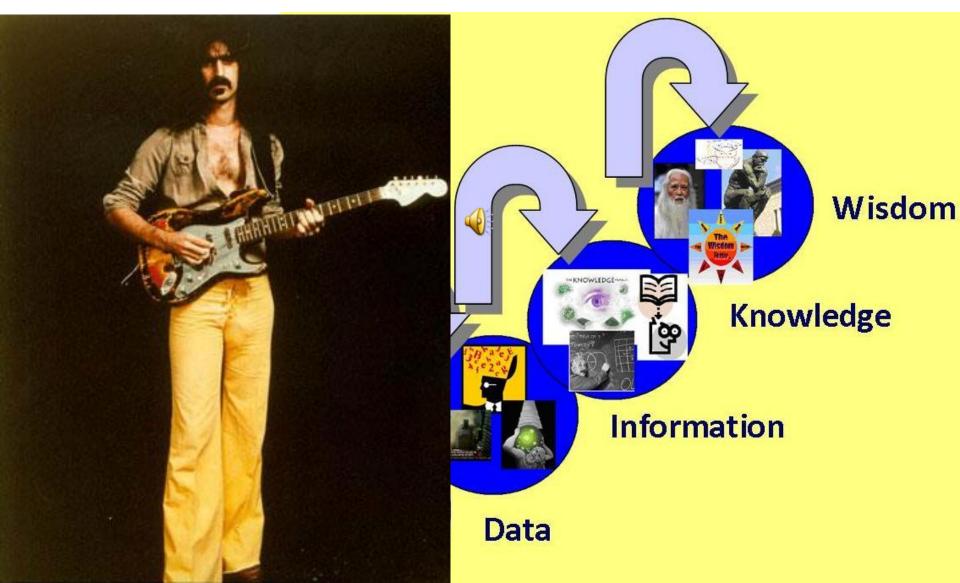
Global IoT Market's Rapid Growth

With the growth of connected devices, the global IoT market size is expected to also flourish. The chart below represents the IoT market forecast from various noteworthy research firms.

IoT Analytics – *Quantifying the connected world*

Global IoT/IoE revenue forecasts

Annual revenue of the IoT market (in Trillion USD)

	Source	Date	CAGR[1] 2014–'20
	IDC	2014	21%
	visiongain	2014	13%
	Harbor Research	2014	33%
	MARKETS AND MARKETS	2014	31%
	Gartner	2013	-

7 Trillion

2.3 Trillion

1.6 Tr

1.0 Tr

800Bn

200Bn

300Bn

© IoT-Analytics.com 2014

2014 2015 2016 2017 2018 2019 2020

Note: Some forecasts only for specific years, in that case all other years in between are extrapolated based on the corresponding growth rate
1. CAGR = Compound annual growth rate 2. Gartner's starting point 2014 not known. Assumption: Same starting point as Markets&Markets
Sources: Gartner, IDC, Harbor Research, Visiongain, Markets&Markets, IoT Analytics

Reference point: Combined revenue of Apple, Google, and Facebook together in 2014

Source : IoT Analytics

SparkLabs
Global Ventures

DIKW Hierarchy

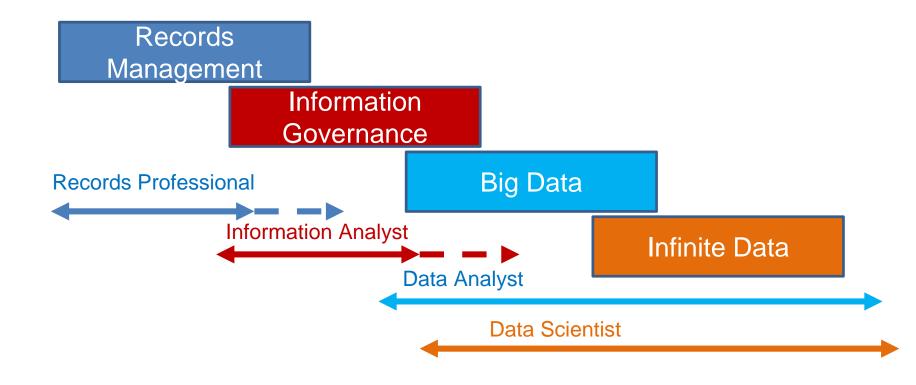# Demanding Customers

Demanding Customers

- Looking for:
  - Answers
  - Analytics
  - Forecasting/Predictions
  - Visualization (coming soon – 3D)
  - From Storage/Retrieval/Governance to Finding Answers

Professional Roadmap

The Role for Government

- Training the next generation of data professionals
  - Data Analytics/Data Science
  - 1m shortfall between now and 2020
- Invest in tools
- Experimentation and risk taking
- Procurement??

The End