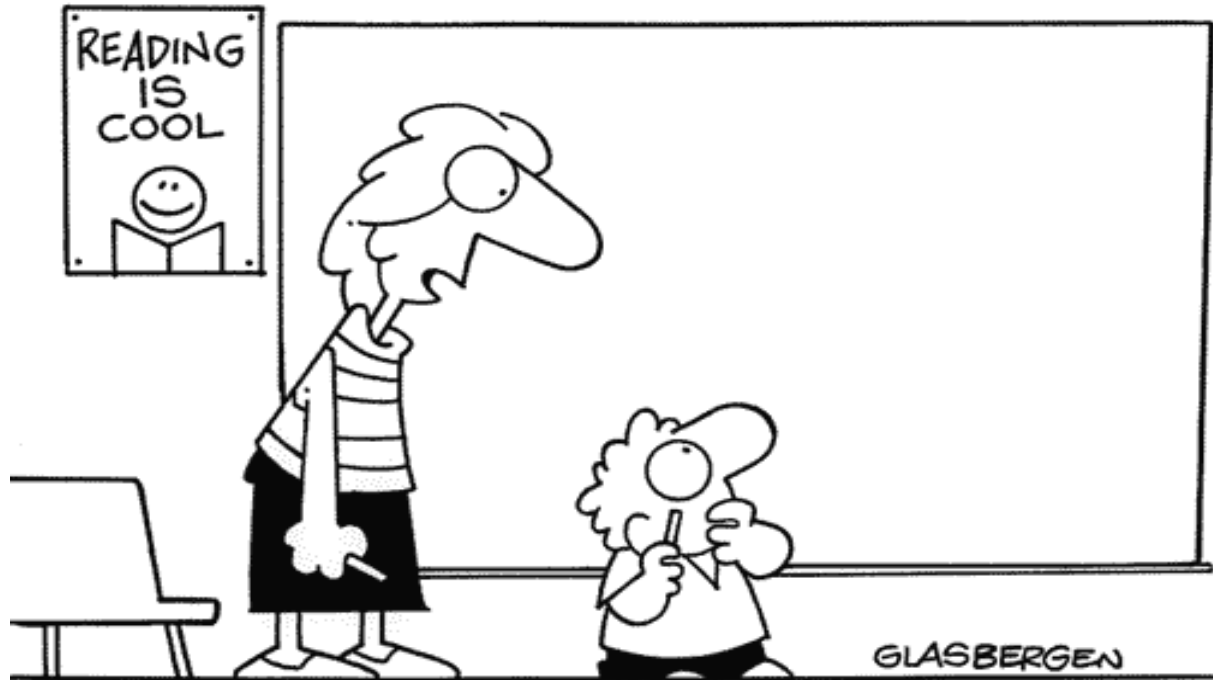


# Mining for Science and Engineering

Presented by: Kenji Yoshigoe

Copyright 1996 Randy Glasbergen. [www.glasbergen.com](http://www.glasbergen.com)



**“There aren’t any icons to click. It’s a chalk board.”**

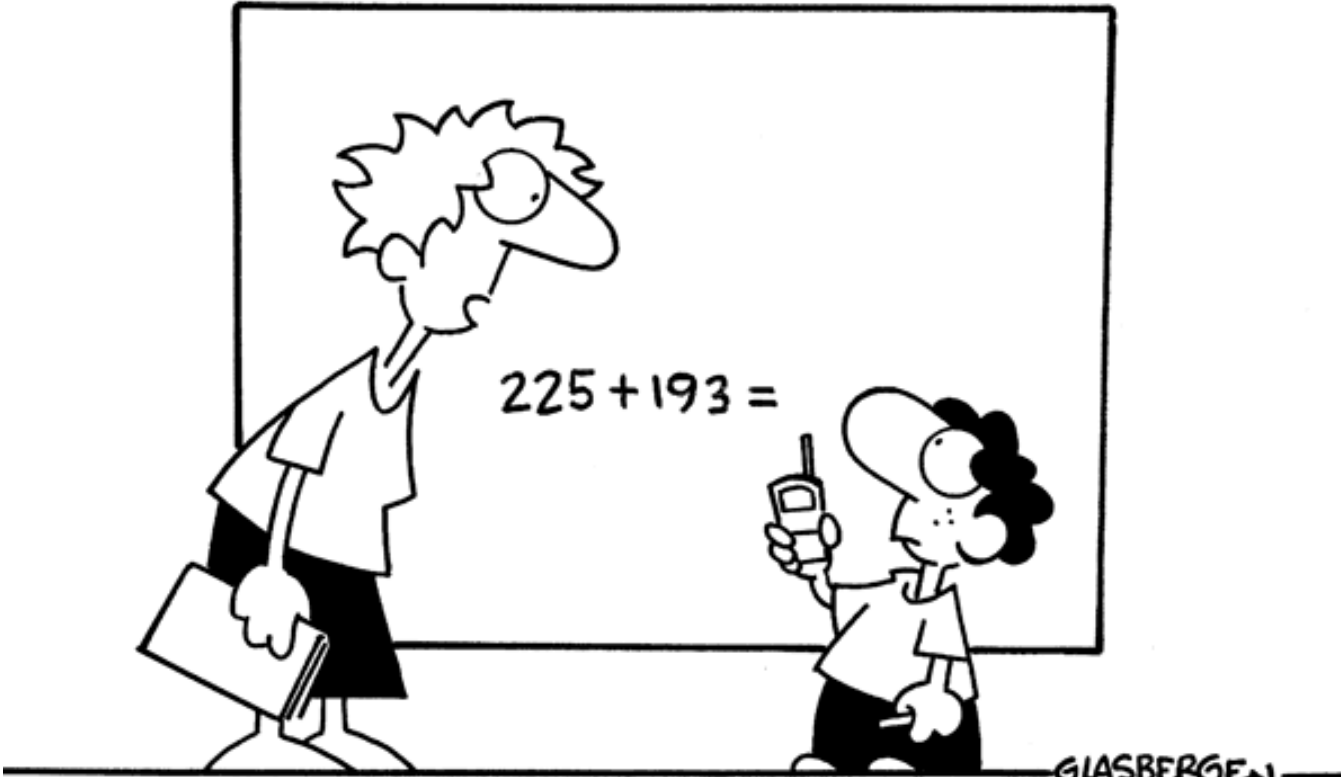
# Why is Data Analytics Critical Now?

- Basis for next 'paradigm-shifting' revolution
  - Need to put the right information in the right person's hand at the right time
  - Impacts: Weather, bioinformatics, biological sciences, emergency management, business intelligence, agricultural sciences, etc.
- Boundary-independent analysis and flow of data requires a flexible, standards-based, 'open' technical (software-driven) infrastructure
  - Design software systems to enable collaborative work
- To facilitate transformative, interdisciplinary research
- Based upon existing strengths and scaling up from current EPSCoR WiNS center effort

# Why is happening now?

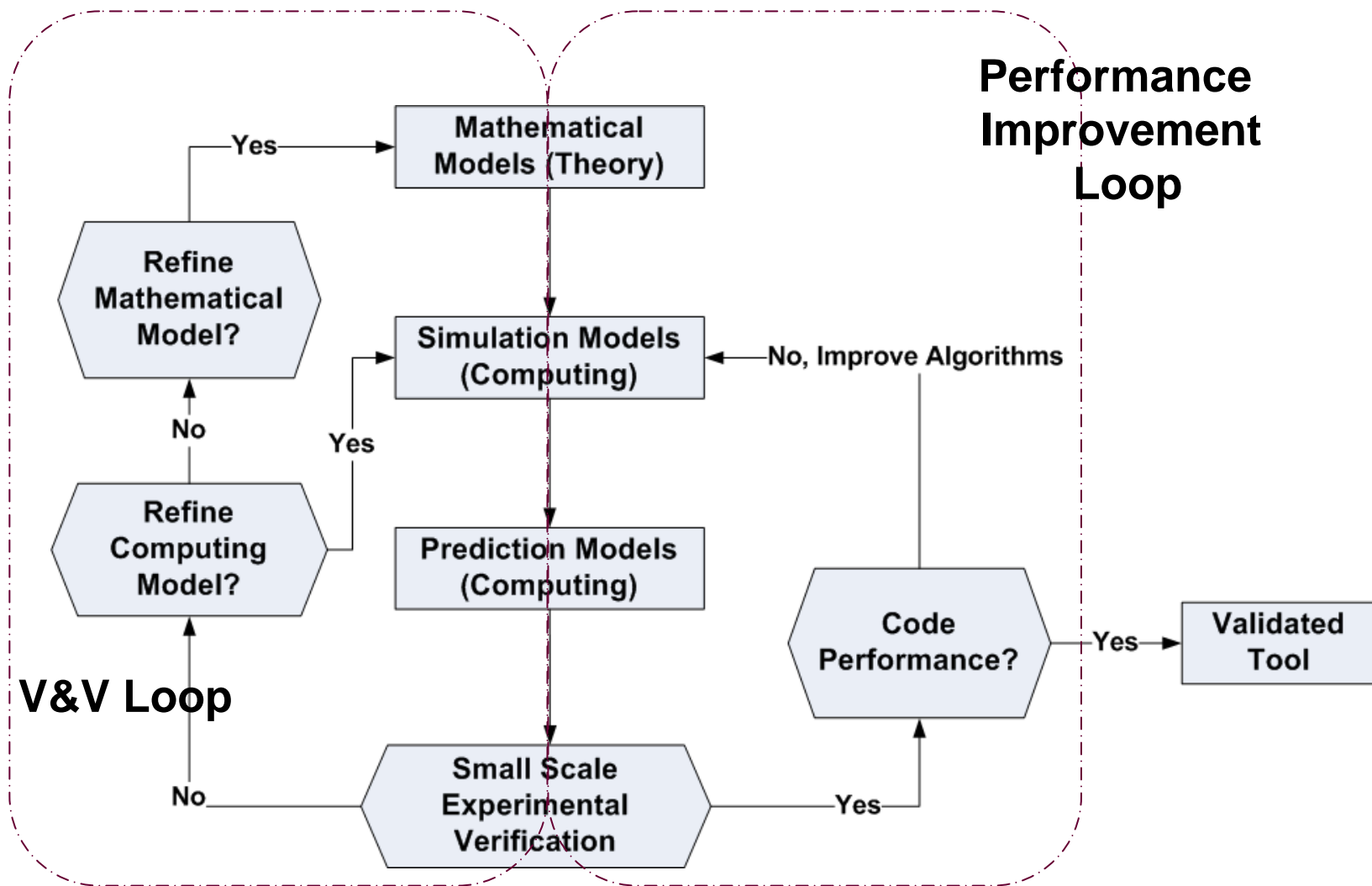
- Estimated that in less than three years, the bytes of data generated by systems and devices will equal the number of grains of sand on all of the world's beaches
- 44 percent of large organizations (i.e., 1,000 employees or more) collect at least 1 terabyte of log file per month
  - 11% collect over 10 TB – per month!
- Security and vulnerability management software products: 20 percent growth in 2007 - ~\$2.27B
  - More than 6 percent to 7 percent growth expected for the software industry as a whole

Copyright 2005 by Randy Glasbergen. www.glasbergen.com



**“You have to solve this problem by yourself. You can’t call tech support.”**

# Data Bottlenecks...

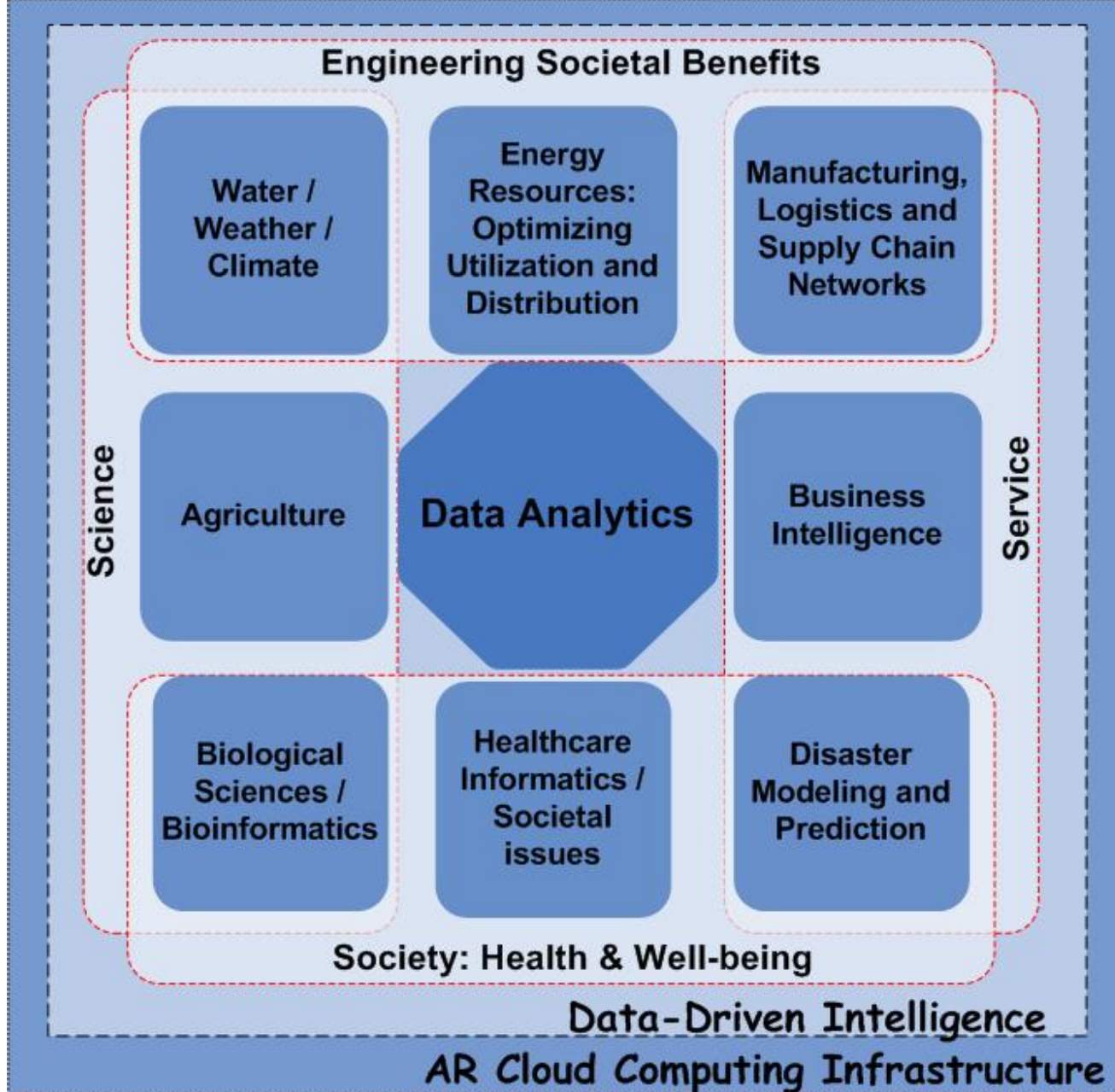


# The Gathering Storm...

- Significant advances in processing, networking and storage have exploded the availability of usable data
  - Need: corresponding advances in analyzing and utilizing gathered data
  - Fundamental advances in how we collect, transmit, protect and ‘process’ valuable information
- Impacts: Weather, bioinformatics, biological sciences, emergency management, business intelligence, agricultural sciences,

# Scientific Objective

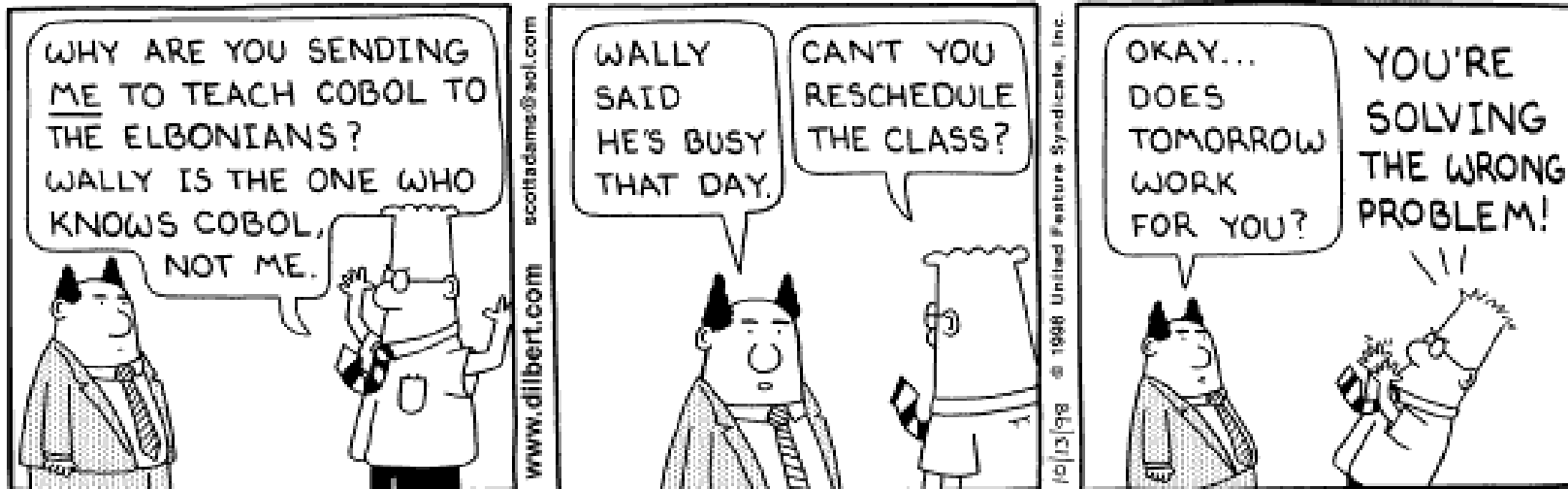
- **Problem:** Data analytics – a fundamental ‘science’ barrier that can enable ‘transformative’ research in multiple disciplines
  - Dealing with data explosion - a critical issue in every domain
- **Aim:** A **transformative infrastructure** for computational data analytics leading to the development of a **computational framework** that is currently lacking
  - **Develop** novel algorithmic analysis strategies for data-driven dynamic systems
  - **Apply** these methods to transform current research across multiple disciplines
  - Data structures capable of handling large, irregular data patterns with historic, time and/or event dependent, evolutionary and semantic characteristics





# Summary

- Strength-based collaborative partnerships
  - UALR, UAF, UAMS, ASU, UAPB
  - Looking for others
- Contact
  - Srini Ramaswamy - [srini@acm.org](mailto:srini@acm.org) / [srini@ieee.org](mailto:srini@ieee.org)



Copyright © 1998 United Feature Syndicate, Inc.  
Redistribution in whole or in part prohibited

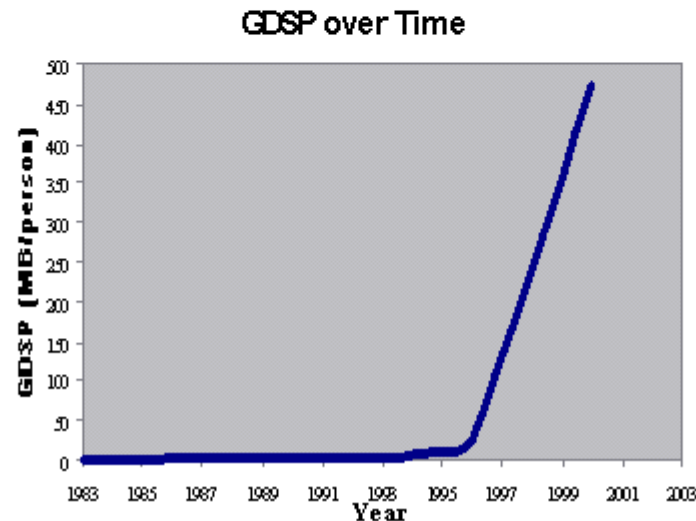
# Why is this important?

- Underlying Issue
  - Dealing with data explosion has become a critical issue in every domain

# Fundamental Barrier Across Science and Technology Disciplines

- Googling for dirt
  - Many profitable new companies focusing on this – H5, Unityware, etc.
  - Law - 500 million documents on each other in response to discovery requests, automated intelligent searches are the only realistic solution

	1983	1996	2000
<b>Storage (TB)</b>	90	160,623	2,829,288
<b>Population(mil)</b>	4,500	5,767	6,000
<b>GDSP (MB/person)</b>	0.02	28	472

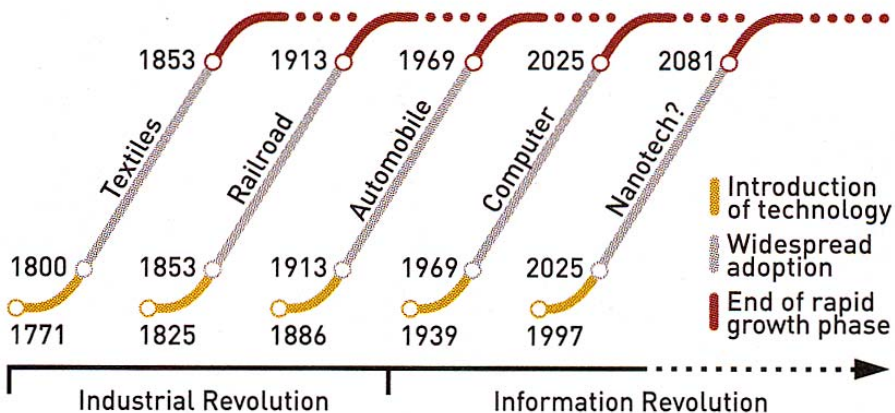


\* Global Disk storage per person GDSP (MB/person)



# REVOLUTIONARY FORCES

Basic advancements in science and technology come about twice a century and lead to massive wealth creation.



SOURCE: Norman Poire, Merrill Lynch

