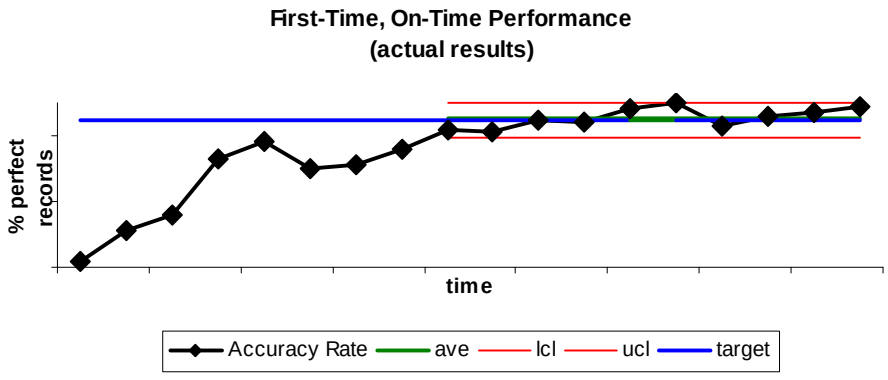
Data Driven: Profiting from Your Most Important Business Asset

Thomas C. Redman, Ph.D.
Navesink Consulting Group

At MIT Conference
Cambridge, MA
July 16, 2008
www.dataqualitysolutions.com

Those who apply diligent efforts (almost) always improve data quality. And benefit!



Each error not made saves an average of $500.

This amounts to millions quickly!

Paradox, part 1

How do we reconcile the evident successes with the observation that data quality is so poor at so many companies?

Hypothesized Answer:

- They (usually) don't recognize data as assets (after all "out-of-sight, out-of-mind) and so worth the effort.

Consultant's exercise: Fire!

You can save only one of the following:

- Antique French Desk.
- Brand new PC, with all the bells and whistles.
- Only copy of the organization's fifty biggest accounts.

The Data Doc's Response: Finally!

The paradox, part 2

How do we reconcile the fact that everyone intuitively knows that data are critical assets with the fact that organizations don't manage them as such?

Hypothesized Answer(s):

- They don't understand what "manage data assets" really means.
- Specifically, they don't see how to make money with data.

What Does "Manage Data Assets" Really Mean?

Generally recognized as business assets:

- Capital, in its various forms
- People, including the knowledge in their heads.

Organizations naturally manage their assets…

- They take care of them.
- They put them to work, to make money.
- They adjust their management systems to account for the special properties of each asset.

For data, "taking care" is mostly about quality

Prescription 1: Take steps to ensure that

- Possess and acquire the right kinds of data.
- People can access and understand them.
- People can trust that they are "good enough."
- They are of high enough quality to withstand market scrutiny.
- They are kept safe from loss or theft.

It is highly significant that (almost) all organizations that diligently follow many of "the ten habits" make order-of-magnitude improvements.

Putting data to work

Prescription 2: Use data to create new revenue

- Sell them directly in the market.
- Build them into other products and services.
- Use them to enhance other products/services.
- Use them to make better decisions.
- Use them to improve the day-in, day-out running of the business.

Critical point: Management must explicitly think through how they will put data to work in creating new value.

Adjusting the management system

Prescription 3: Recognize that data have unique properties

- Example: Unlike other assets, data can be shared
- Most important: Data are the only asset that are uniquely an organization's own. The "ultimate proprietary technology."

Prescription 3, cont: Adjust the organizational structures, roles, and responsibilities as a result.

- Counterexample: Chief Information *Technology* Officer

Outline:

- What does “manage data assets” mean?
- A bit of flavor for:
  - Putting data to work
  - The wondrous and perilous properties of data as an asset
  - Implications for the management system
  - The brutal (and growing) politics associated with data
- A new context for data quality
- The ten habits of those with the best data.

Putting Data to Work

- The many ways to bring data to market
- A note on decision-making
- Your most important data

A Note on Market Demands

- People and organizations have always wanted "more and better" data.
- Historically, the elite took steps to hoard data.
- Since the rise of democracy, some of their grip has been broken.
- Sheer demand continues to grow and is in little doubt:

"Inside IBM, we talk about 10 times more connected people, 100 times more network speed, 1,000 times more devices, and a million times more data."*

*Lou Gerstner, quoted in McDougall, P., "More Work Ahead," *Information Week*, December 18-25, 2000, p. 22.

A Note on Market Demands-2

- To borrow from Twain,

  "the difference between the right data and the almost-right data is like the difference between lightning and a lightning bug."

- People and organizations expect:

  "exactly the right data and information in exactly the right place at exactly the right time and in the right format to make a decision, complete an operation, or serve a customer."

So far, I've identified fifteen ways to fulfill these demands

Provide Content

- New Content
- Re-package
- Informationalization
- Unbundling
- Exploiting Asymmetries
- Closing Asymmetries

Facilitators

- Own the Identifiers
- Infomediation
- Data mining/Analytics
- Privacy and security
- Training
- New Marketplaces
- Infrastructure technologies
- Information appliances
- Tools

Content Providers

Basic Idea: Provide newer, richer, better, etc. data to address customer needs

Customer Need Fulfilled: "Which diet will work best for me?/What can I realistically expect to achieve?"

Industrial Age Examples: Generic diet guides, Newspaper

Information Age Examples: Personal diet, Morningstar, Information Resources Inc.

Informationalization

Basic Idea: Enhance existing products and services by building data and information in.

Customer Needs Fulfilled: Simplicity, integration, etc.

Industrial Age example: Product instructions

Information Age examples:

- Auto makers are now including GPS Navigation systems.
- NC State is re-designing the hospital gown, building a thermometer and other sensors in.

Exploit Information Asymmetries

Basic Idea: Know more than the guy on the other side of the transaction.

Customer Need Fulfilled: Get the best possible "deal."

Industrial Age example: Used car salesman

Information Age examples: Hedge Funds

Infomediation

Basic Idea: Help people find the data and information they need

Customer Need Fulfilled: Waste less time.

Industrial Age example: Travel agents

Information Age example: Google

Data Mining/Analytics

Basic Idea: Uncover hidden "nuggets" buried in the data.

Customer Need Fulfilled (examples):

- Deep insights into individual needs
- Exploit patterns of excellence/opportunities of improvement.

Industrial Age example: Statistical analysis

Information Age example: Harrah's, Amazon

Other Ways to Bring Data and Information to Market - Content

- Repackage and filter to meet specific needs
  - FTID,
- Unbundle
  - Securities Research and trade processing
- Close Information Asymmetries
  - Consumer Reports

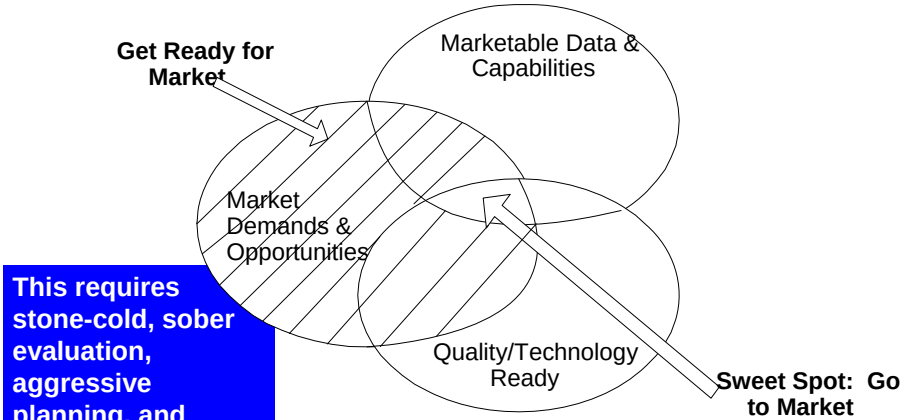
Other Ways to Bring Data and Information to Market - Facilitators

- Own the identifiers
  - Standard & Poors
- Privacy and Security
  - Legal profession
- Define and Operate “data markets”
  - E-Bay
  - Facebook
- Training and Education
  - Internet-based training

Make Better Decisions

Another good way to put data to work is systematically use them to make better decisions, align the organization to the tasks at hand, and execute.

Implication: Organizations need to find and pursue their "data sweet spots"



This requires stone-cold, sober evaluation, aggressive planning, and commitment to innovation!

Implications, cont:

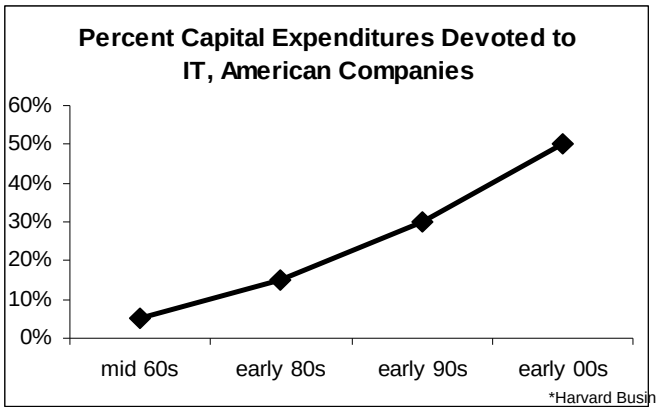
- Data Doc claim: "The organization's most important data are those that help it make money."
- Those used to create new revenue are especially important.
- Note that every organization exposes some data in its marketplaces.
- We data geeks should focus on these business opportunities and the required data.
- We should measure success by metrics like "new revenue from data."
- Note: It is a lot easier to invest in revenue growth than cost reduction. Improved quality is a perfect example.

The Wondrous, Perilous and Often Confounding Properties of Data In Organizations

- Most important: The ultimate proprietary technology.
- Data are "organic."
- Note: About ten such properties really matter.

"IT Doesn't Matter," Nicholas Carr*

Information Technologies have penetrated every aspect of modern life.



*Harvard Business Review, May 2003

Proprietary vs. Infrastructure Technologies

PROPRIETARY	INFRASTRUCTURE
Can be "owned" by a single organization	(Eventually) part of general business infrastructure
Patented drug, unique process	Railroads, electric grid
Protected	Become commoditized
Basis for sustained advantage	Not a basis for sustained advantage

Advantage Stems from Scarcity…

- Carr argues that basic storage, processing, and transport technologies are now readily available to all.
- Carr does not argue that IT isn't important. Only that it is not strategic.
- He offers the following advice:
  - Spend less.
  - Follow, don't lead.
  - Focus on vulnerabilities, not opportunities.

Finding Reasons to Attack Carr is Easy

- No proprietary technology/advantage lasts forever… or even very long.
- The pace of innovation in IT is only growing.
- Advantage can still be sustained by using IT in smarter ways.

But many organizations seem to be following his advice!

Data are the Organization's Ultimate Proprietary Technology!

- No other organization has, or can have, the same data.
- Data are subtle and nuanced.
  - Model "customer" in unique ways that best suit it.
  - Capture and utilize unique "facts."
  - Processes to capture unique data are also difficult to copy.
- Eventually, of course, some data become standardized to facilitate communications.
- Data offer opportunity for sustained advantage—and everyone knows it!

Data are subtle and nuanced and have become the organization's *lingua franca*

USA
Football

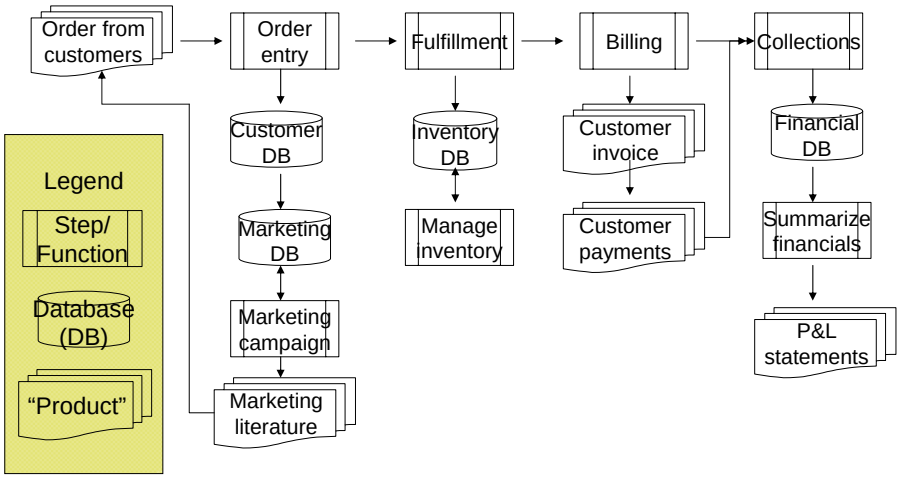
UK
Football

AUSTRALIA
Football

Never underestimate the importance of local knowledge.

HSBC
The world's local bank

Data are "organic" and most useful "on the fly" (when they cross departmental boundaries)



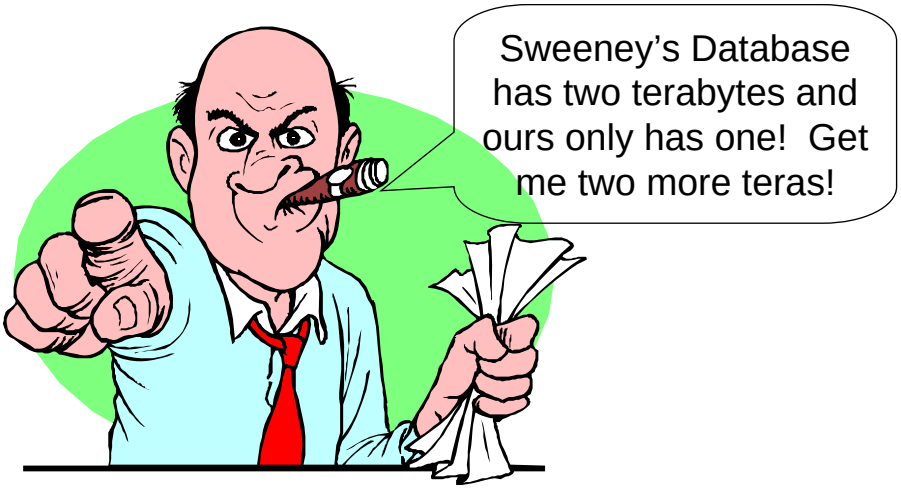
Implications

- Must not confuse management of *technology* with management of *data*.
- Must be very careful about what data we standardize. Standard data has little marketplace value.
- Should strive for greater uniqueness, novelty, and depth in data put in the marketplace.
- Need to identify and explicitly manage the most important, end-to-end value-creating flows of data as "information chains" or Big-P processes.
- Need to improve quality, in its own right, but more especially to meet market demands.

The Surprisingly Brutal Politics Associated with Data

- Data Sharing
- Responsibility for quality
- About a dozen important, as they play out locally.

1. Power/Data Sharing/Ownership: In the Information Age, Possession of Data Conveys Power!

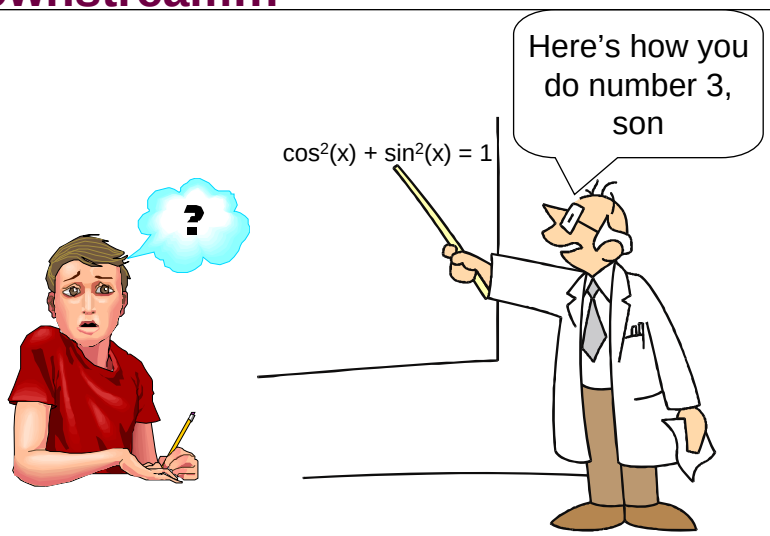


1, cont. Though Universally Praised, Data Sharing is the Exception!

NOTE: Many of *The 48 Laws of Power* (Greene and Elffers, Viking, 1998) seem to argue against sharing data.

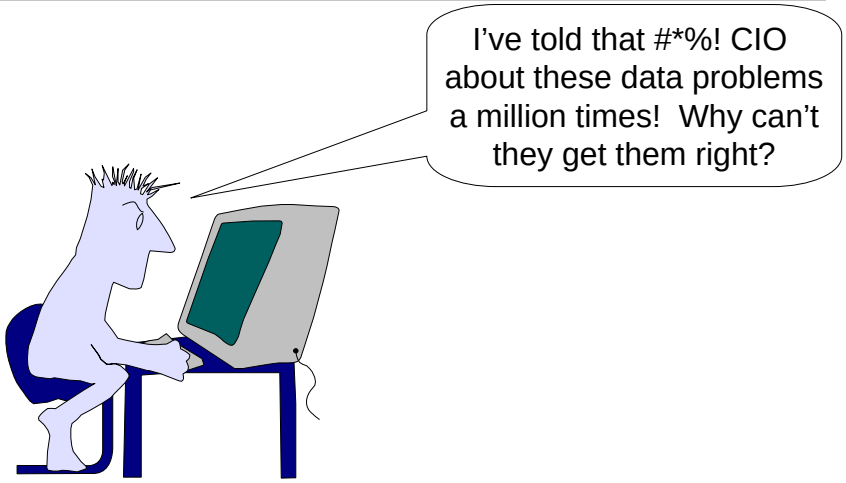
Redman-MIT08-Data Driven © Navesink Consulting Group, 2000-2008 T. C. Redman, Page 37

It is so easy for accountability to shift downstream!!!



Redman-MIT08-Data Driven © Navesink Consulting Group, 2000-2008 T. C. Redman, Page 38

Who is responsible for data quality? Since the data are "in the warehouse," it must be the CIO!



Landauer, *The Trouble with Computers* - 1993

Motivation: Roach, Strassman (early 90s): "Why aren't computers improving productivity?"

- Computers are remarkably effective at "computing:"
  - Switching phone calls
  - Laser-guided weapons
  - Weather forecasting
- Computers are not so effective when automating poorly-defined processes:
  - Word Processing
  - Poorly-defined Business Processes

Landauer, *The Trouble with Computers* - 1993, con't

In some cases, productivity doesn't improve, but there are other benefits:

- ATMs: Not cheaper, but always available.

Landauer's results are consistent with other results:

- Deming: "If you automate a factory that produces junk, you'll just produce junk faster."
- Data warehouses: Add little value unless decision processes are well defined.
- Enterprise Systems: Not accepted unless they match the way people work.
  - Example: $170M Failure in FBI's "Virtual Case File."

Implications

- You can't resolve the inter-related issues of ownership, management accountability, and quality through automation.
- Process management and improvement for quality and effectiveness.
- Automation for speed, efficiency, and scale.
- Need to explicitly get responsibility for data out of the CITO.
- New organization in "the business:" Chief Data Office.
- Those interested in data must learn how to build and use "political capital."
  - Politics" is increasingly important.
  - Note: Politics is NOT inherently negative.

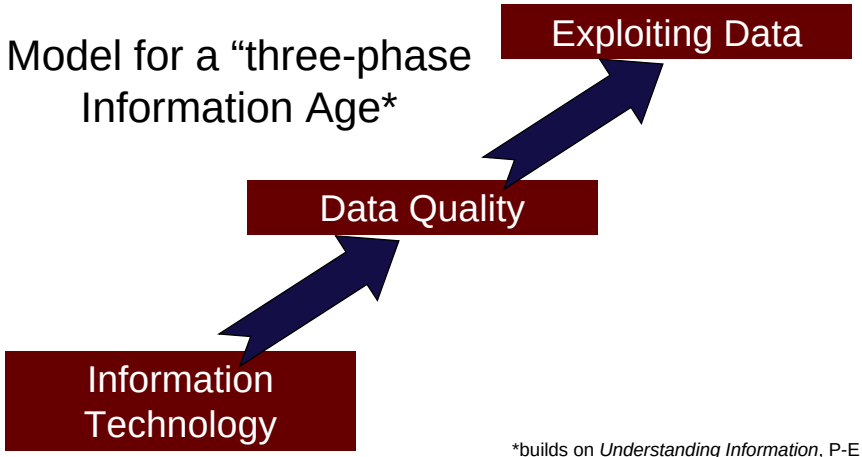
Data Quality

A new and better context

The ten habits of those with the best data

IT set the stage for, and is now giving way to, data

Model for a "three-phase Information Age*



Carr: Nearing completion

*builds on *Understanding Information*, P-E Centre for Management Research, 1994.

Proper Context for Data Quality

Existing	Proper
◻ Cost Reduction	◻ Revenue growth
◻ Internal customers	◻ Competitive advantage
◻ "Master Data"	◻ Data exposed in market
◻ CITO	◻ "The business"
◻ Standardization	◻ Novelty, uniqueness
	The case for data quality (anything for that matter) is a lot easier when it involves new revenue

Those with the best data think "prevention"

Approaches to Data Quality: Defect Prevention

Most companies' current approach to data quality. Typical error rates are 1-5% and "cost of poor data quality" may be 20% of revenue.

THIRD-GENERATION:* Design, defects "impossible"

Don't know of anyone here

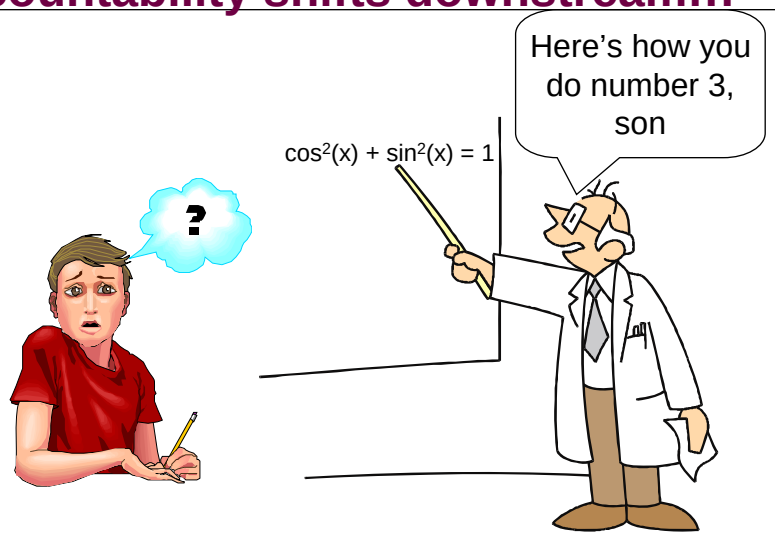
SECOND-GENERATION:* Process/Supplier Management, to prevent defects

FIRST-GENERATION:* Inspection and Rework, to find and fix defects

To accomplish this, original sources of data are held accountable. Typical error rates are 1-2 orders of magnitude better and the cost of poor data quality is reduced about two-thirds.

*terms after Ishikawa

They recognize that, left alone, accountability shifts downstream!!!



The (nearly-certain) results

Approach	Management Focus	Typical Error Rate	Cost of Poor Data Quality
Find and Fix (First-Gen)	The Past	1-5% (at the field level)	20% of revenue
Prevent Future Errors (Sec-Gen)	The Future	Two orders of magnitude better	Reduced by two-thirds

Habit 1: Focus on the most important needs of the most important customers

Those with the best data adopt a customer-facing definition of quality.

In doing so, they recognize that:

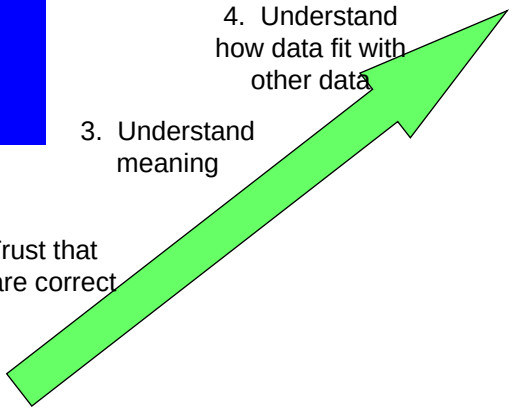
- All data are not created equal. Similarly, customers, problems, and business opportunities are not created equal.
- Generally, the most important data are those needed to set and execute the company's most important business strategies.

And they focus as much of their energies on these customers, strategies, and data.

Said differently, their data quality programs are fully aligned with business strategy.

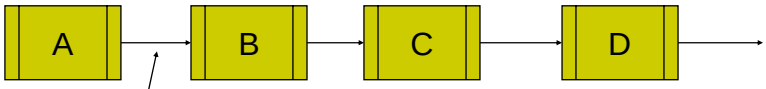
Data Doc's Hierarchy of Needs

Many people and organizations exhibit a "hierarchy of "needs"



Habit 2. Process, process, process

They recognize that they create data via their cross-functional business processes

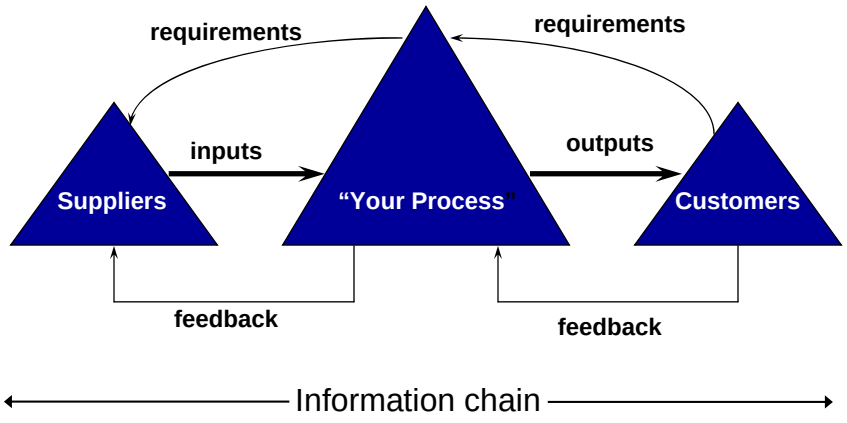


They recognize that most errors occur "in the white space

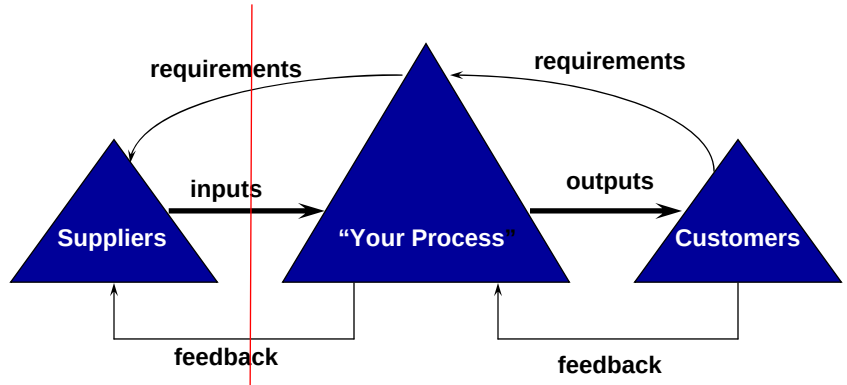
They think "BIG-P"

They recognize "the next guy" (serving the customer) as a customer

They use the Customer-Supplier Model to establish requirements and feedback loops



Habit 3: They employ supplier management for external sources of data



They expect high-quality data from outside. And invest (time) with their suppliers to get them

Habit 4: They measure quality at the source in business terms

They define metrics with clear business implications.

Private Bank's Customer Data:

Percent of statements with an error

Telecom's Access Charges:

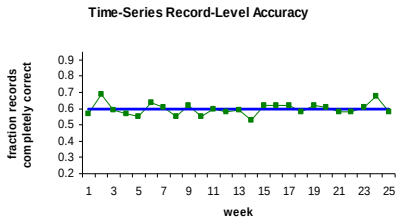
Risk = Overbilling + Underbilling

Many organizations:

Fraction "perfect" records

(interpreted as "work" done correctly)

They measure continuously



They get good at interpreting results

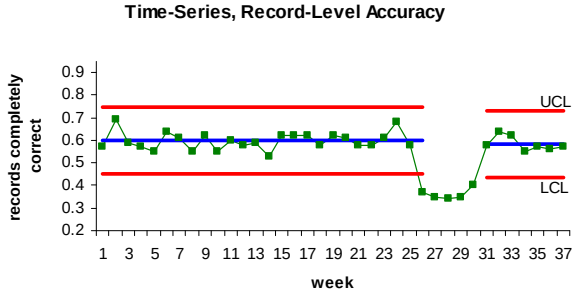
They integrate top-line DQ metrics with other business results

Habit 5: They employ controls at all levels to halt simple errors and establish a basis for moving forward

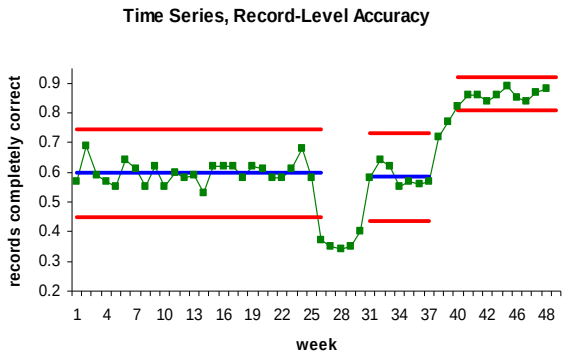
They employ simple edits to stop errors in their tracks:

Ex: (Title = Mrs., Sex = M) cannot be correct

They employ statistical control to identify process issues early and to look forward:



Habit 6: They have a knack for continuous improvement



They have a way of not just starting, but completing improvement projects, both to:

- eliminate root causes of error
- acquire new data

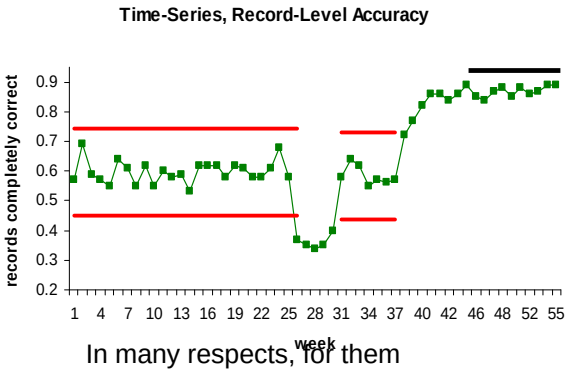
Habit 7: Set and achieve aggressive targets

They focus not just on the level, but also on the rate of improvement

They set targets like:

- half the error rate every year
- add two significant new features every year

They decide to position themselves near the front with respect to quality in their industries



In many respects, for them planning for quality is no different than planning for revenue growth, new product development, etc.

Habit 8: Formalize management accountabilities for data

They recognize that responsibility for data lies with "the business," not IT.

Some codify responsibilities in policy.

My favorite (adopted for data):

"Don't take junk data from the guy upstream. And don't pass junk data on to the next guy!"

Redman-MIT08-Data Driven © Navesink Consulting Group, 2000-2008 T. C. Redman, Page 59

Habit 9: A broad, senior group leads the effort

- They know that that quality programs go as far and fast as the senior person leading the effort demands.
- So a broad, committed, senior team leads the effort.

"They thought they could make the right speeches, establish broad goals, and leave everything else to subordinates... They didn't realize that fixing quality meant fixing whole companies, a task that can't be delegated."

Dr. Juran, 1993

Experience so far is that "data" is even tougher than the factory floor.

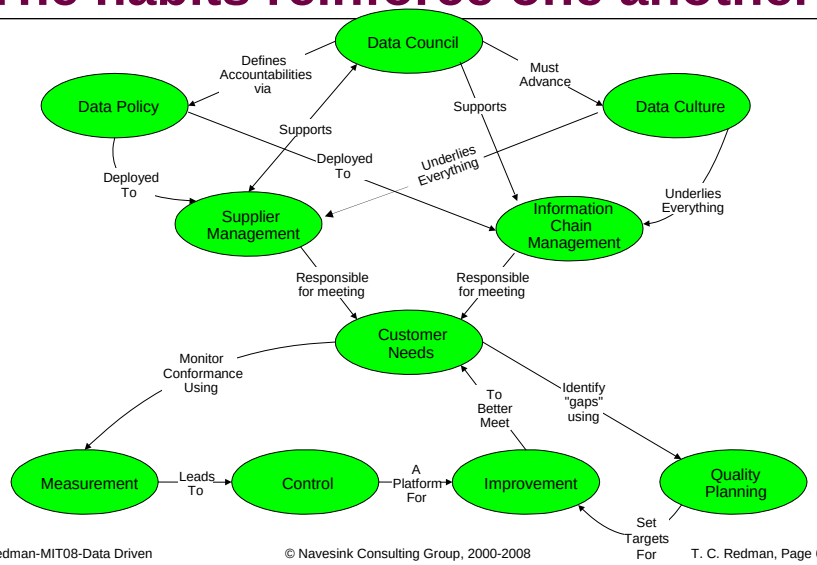
Redman-MIT08-Data Driven © Navesink Consulting Group, 2000-2008 T. C. Redman, Page 60

Habit 10: Recognize that the "hard issues are soft" and actively manage change

They:

- Distinguish "I" from "IT." They recognize that they can't automate their way out of a quality issue.
- Start small. Create early wins.
- Actively manage change.
- Avoid unwinnable battles, especially early on.
- Build political capital.
- Over time, they build data quality into:
  - The organization
  - People's psyche
  - To new systems

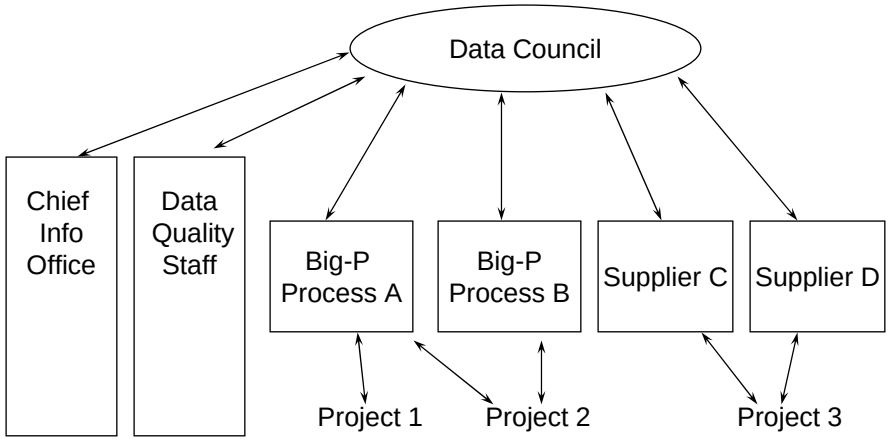
The habits reinforce one another



The Ten Habits apply to all data, in all industries and government

- Market, product, and people (customer and employee) data. Intelligence, scientific and logistics data. Health care data.
- Data created internally or gathered from external sources.
- Meta-data, master data, enterprise data.
- Data to be stored on paper, in operational systems, in warehouses, enterprise systems.
- Client statements, 10-Ks, prospectuses.
- Data only seen by computers and data that convince people to trust industries and companies (or not).

Proposed Organizational Model for Data Quality*



*overlaid on current organization

Final Remarks:

"Data are assets" and they deserve to be managed as professionally and aggressively as other assets.

- Put them to work, especially in the market.
- Recognize that they are unlike other assets and advance the management system to account for, and leverage, these differences.
- From a quality perspective, the rigors of the marketplace should drive quality requirements.
- Follow the ten habits to meet marketplace requirements.

What Did He Say?

Thomas C. Redman, Ph.D.
Navesink Consulting Group
President
732-933-4669
www.dataqualitysolutions.com

