Correcting the Misconceptions About the Nature of Data that Thwart Information Quality

ABSTRACT

Understanding the true nature of data has enabled Fed Ex and UPS to track packages across their disparate data silos, the CIA to "connect the dots" in the war on terror, and Bechtel to revolutionize their IT organization.

Data is more than the byproduct of computerized processes, and more than the property of applications. And, as the high failure rate of business intelligence projects indicates, data is more than a fungible commodity that can be easily replicated, aggregated, sliced and diced. When IT professionals harbor misconceptions about the nature of data, information quality becomes a chronic problem they can never solve.

When IT professionals understand the true nature of data, when they think of data as a record of reality, then business and IT professionals can use data to evaluate systemic risk on a macro level, manage individual assets on the micro level, and spot hidden opportunities for the future.

BIOGRAPHY

Lyn Robison
Research Director
Burton Group

Lyn Robison is the Research Director for Burton Group's Data Management Strategies service. Lyn directs the research of Burton Group analysts who provide guidance for Global 2000 corporations on data management issues. He is the author of two books and numerous articles on information technology, and is a frequent speaker at IT industry conferences.

Lyn Robison
Research Director
Burton Group
Data Management
Strategies
LRobison@burtongroup.com
www.burtongroup.com

Correcting the Misconceptions about the Nature of Data that thwart Information Quality

MIT 2010 IQIS
Friday – July 16, 2010

Misconceptions about the Nature of Data

- **Thesis**
  - Data is more than the byproduct of computerized processes, more than the property of software applications, and more than a fungible commodity that can be easily replicated, aggregated, sliced and diced.
  - When IT professionals harbor misconceptions about the nature of data, information quality becomes a chronic problem they can never solve.
  - When IT professionals understand the true nature of data, when they think of data as a description of reality, then business and IT professionals can use data to evaluate systemic risk on a macro level, manage individual assets on the micro level, and spot hidden opportunities for the future.

Misconceptions about the Nature of Data

- **Agenda**
  - Misconception: Data is the byproduct of processes
  - Misconception: Fungibility is No Big Deal
  - Misconceptions about data make IQ a chronic problem
  - Treating data as a picture of reality leads to high IQ
  - Recommendations

Misconceptions about the Nature of Data

- **Agenda**
  - ***Misconception: Data is the byproduct of processes***
  - Misconception: Fungibility is No Big Deal
  - Misconceptions about data make IQ a chronic problem
  - Treating data as a picture of reality leads to high IQ
  - Recommendations

Misconception: Byproduct of Processes

burton GROUP™

- **Misconception: "IT's job is to automate business processes!"**
  - Data is viewed only in the context of the application or the system or the business process that created it
  - Data is viewed merely as the input/output for business processes
    - If a process needs some data as input, it is perfectly okay to blithely replicate some data so that the process can get the data input it needs
    - The processes are the priority, and the data is simply the juice to get the processes started and merely the byproduct after the processes are completed
  - A current, real-life example of this mindset is SOAP
    - SOAP is designed for feeding input to and for obtaining output from processes

Misconception: Byproduct of Processes

burton GROUP™

- **Misconception: "IT's job is to automate business processes!"**
  - Some IT professionals view data as sawdust – merely the byproduct of the process

Misconception: Byproduct of Processes

- **Misconception: "IT's job is to automate business processes!"**
  - This concept is a hold-over from the mainframe batch processing days
  - Computer systems were not designed to preserve data, ~~~~~~~te

http://www.flickr.com/photos/carrick/61952845/

Misconception: Byproduct of Processes

- **Correcting this misconception**
  - Modern computers can actually store data
    - Computers are no longer punch card processing mainframes
  - Of course, a big part of IT's job is to automate processes
    - The business expects enterprise IT to automate their work
  - However, the business needs more from enterprise IT than process automation
    - Information, if it is properly preserved and managed, can be a highly valuable corporate asset
    - Information systems can and should let people collaborate and build relationships with one another

Misconceptions about the Nature of Data

- **Agenda**
    - Misconception: Data is the byproduct of processes
    - ***Misconception: Fungibility is No Big Deal***
    - Misconceptions about data make IQ a chronic problem
    - Treating data as a picture of reality leads to high IQ
    - Recommendations

Misconception: Fungible Commodity

- **Misconception: "Data is merely a fungible commodity!"**
    - Gasoline is a fungible commodity

Misconception: Fungible Commodity

- **Misconception: "Data is a commodity like oil!"**
  - "We can fill up databases with data the same way that steel barrels can be filled with crude oil."

Misconception: Fungible Commodity

- **Misconception: "Data is merely a fungible commodity!"**
  - People who hold to this misconception believe that, even though computer systems can be designed to preserve data, all data is merely a fungible commodity that can be easily aggregated
  - Databases are viewed as containers that hold data the way that steel barrels hold crude oil or railroad cars hold lumber
  - IT professionals create databases and then the business processes fill 'em up with data
  - Data warehouses and BI tools are used to aggregate this data
  - Data produced through this perspective makes it possible for corporations to manage fungible assets such as money
    - Numbers of dollars for revenue, expenses, assets, and liabilities are totaled up for each division

Misconceptions about the Nature of Data

- **In fact, in the case of money, there is no need to keep track of individual dollars, because dollars are fungible**
  - One dollar is just as good as another and all that is needed are the totals

Misconception: Fungible Commodity

- **Misconception: "Data is a commodity like cut lumber!"**
  - Misconception: all data can be easily aggregated into

Misconception: Fungible Commodity

- **Misconception: "Data is a commodity like cut lumber!"**
  - Misconception: all data can be easily moved around using ETL

Misconception: Fungible Commodity

- **Misconception: "Data is a commodity like cut lumber!"**
  - Misconception: all data can be delivered up to

Misconception: Fungible Commodity

- **Misconception: "Data is a commodity that we can slice and dice!"**
  - Misconception: all data can be sliced and diced like vegetables

Misconception: Fungible Commodity

- **Correcting these misconceptions**
  - In information systems, long-lived, non-fungible assets must be represented differently than fungible commodities
    - "Type" is not the only consideration
    - "Identity" is vital to capture properly and manage beyond silos
  - The fact is that unless computer systems are designed to preserve unique instance data and indentify individual, non-fungible assets across silos, those computer systems are unable to manage non-fungible assets such as customers
    - Hence we see the huge amounts of money spent on ineffective CRM systems over the past couple of decades
    - Dedicated CRM silos don't succeed
    - Distributed customer data can't be easily brought together

Misconceptions about the Nature of Data

- **Agenda**
  - Misconception: Data is the byproduct of processes
  - Misconception: Fungibility is No Big Deal
  - ***Misconceptions about data make IQ a chronic problem***
  - Treating data as a picture of reality leads to high IQ
  - Recommendations

IQ is a Chronic Problem

- **Fungible or Non-fungible?**
  - Is currency fungible? Usually.

IQ is a Chronic Problem

- **Fungible or Non-fungible?**
  - Are these airplanes fungible? It depends.

IQ is a Chronic Problem

- **Fungible or Non-fungible?**
  - Are buildings fungible? Not usually.

IQ is a Chronic Problem

- **Fungible or Non-fungible?**
  - Are people fungible? No.

IQ is a Chronic Problem

- **Fungible or Non-fungible?**
  - Are contracts and agreements and relationships fungible? No.

IQ is a Chronic Problem

- **What happens when information systems use techniques designed for representing fungible commodities to represent non-fungible assets?**
  - Distortion

IQ is a Chronic Problem

- **What happens when information systems use techniques designed for representing fungible commodities to represent non-fungible assets?**
  - Incomplete information

Misconceptions about the Nature of Data

- **Agenda**
  - Misconception: Data is the byproduct of processes
  - Misconception: Fungibility is No Big Deal
  - Misconceptions about data make IQ a chronic problem
  - ***Treating data as a picture of reality leads to high IQ***
  - Recommendations

Treating Data as a Picture of Reality

- **Data is not a computer phenomenon**
  - Data existed long before computers were imagined
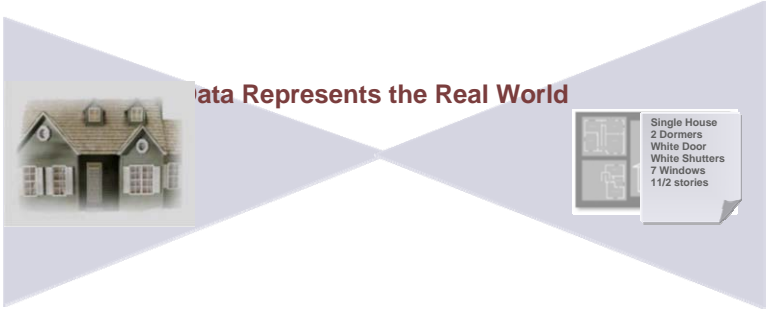 - Data is a way to augment humans' limited capacity to remember

Treating Data as a Picture of Reality

- **Data is a description of relevant portions of reality**
  - Data is used to manage current reality, record/look at the past, and predict the future
  - The validity and usefulness of data depends directly on how accurately it represents real world

ata Represents the Real World

Single House
2 Dormers
White Door
White Shutters
7 Windows
11/2 stories

Treating Data as a Picture of Reality

- **The world is infinitely complex**
  - Data represents those aspects of reality that are important to us
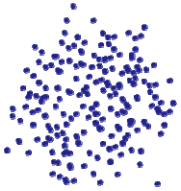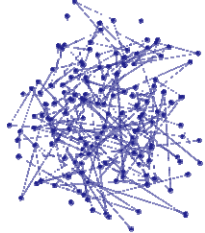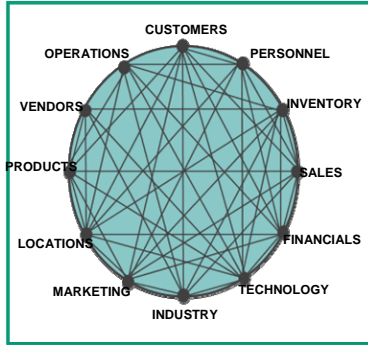
Treating Data as a Picture of Reality

- **High IQ means data that represents relationships**
  - Particularly the relationships between non-fungible assets

Individual data items

The interdependency of data

Treating Data as a Picture of Reality

- **Managing data about non-fungible assets is vital**
  - Silos make this difficult
  - Silos are a permanent fixture of enterprise IT

Misconceptions about the Nature of Data

- **Agenda**
  - Misconception: Data is the byproduct of processes
  - Misconception: Fungibility is No Big Deal
  - Misconceptions about data make IQ a chronic problem
  - Treating data as a picture of reality leads to high IQ
  - ***Recommendations***

Recommendations

- **Make a deliberate study of fungibility in your information systems**
  - Make a clear distinction between your fungible commodities and your non-fungible assets
  - When dealing with data representing fungible commodities, you can concern yourself primarily with data "types"
  - When dealing with data representing non-fungible assets, you must concern yourself with data "types" *and* "instances"

Recommendations

- **Implement your information systems based on fungibility**
  - Silos are not a big problem for fungible commodities
    - Data warehouses, ETL, and traditional BI applications work fine to aggregate, analyze, and report on fungible commodities
    - Data modeling is vital to define your data types clearly
    - See Burton Group's free guidance on data modeling:
    - Data Modeling—A Necessary and Rewarding Aspect of Data Management at http://www.burtongroup.com/Guest/Dm/DataModelingANecessary.aspx
  - Silos are a huge problem for non-fungible assets
    - Don't try to eliminate silos – they are permanent
    - Use silo bridging instead of silo busting
    - Manage the identifiers of non-fungible assets beyond silos
    - See Burton Group's free guidance on MODS: Delivering Integrated Information from Data Silos Using MODS at http://www.burtongroup.com/Guest/Dm/DeliveringIntegrated.as

Recommendations

- **Take periodic, regular measurements of IQ**
  - Track data management maturity and information quality consistently over time
    - Track data management maturity and conduct user surveys of information quality on a periodic, ongoing basis
    - Measure information quality and data management maturity before and after each relevant IT project and track and report on the improvement or lack of improvement
  - See the 2009 MIT IQIS presentation entitled, "IQ Metrics: Diagnosing IT's Impact on the Business"

Recommendations

- **Build information systems that deliver highly useful information**
    - Evaluate systemic risk on a macro level
    - Manage individual assets on the micro level

Recommendations

- **Build information systems that treat data according to the types of things that the data represents**
    - Spot hidden opportunities for the future