Data Quality – Where did it all go wrong?

Ed Wrazen, Trillium Software
• Examples of data quality problems
• Why do data quality problems occur?
• The impact of poor data
• Why data quality is an enterprise issue
• Mitigating the risk of poor data quality
Examples of Data Quality Problem

- Retail company found over 1m records contained home tel number of “0000000000” and addresses containing flight numbers
- Insurance company found customer records with 99/99/99 in creation date field of policy
- Car rental company discovered duplicate agreement numbers in their European data warehouse
- Healthcare company found 9 different values in gender field
- Food/Beverage retail chain found the same product was their No 1 and No 2 best sellers across their business
Real World Examples

Order this: a part for the Sea Harrier weapons system

Get this: A plastic bag
Real World Examples

Order this:
1 day ration pack

Get this:
even the Marines can’t eat valves!
Real World Examples

One system’s view

Defence Catering Group

Stock Number: 99 000 1111
24 Hour Cold Climate Ration Pack

Another system’s view

Army Stores System 3

Stock Number: 99 000 1111
Valve Electronic
Data Impacts Financial Results

- Drive revenue
- Reduce operational costs
- Optimize processes
- Create growth opportunities

Data is an important corporate asset that drives business processes, strategic decisions and financial results.
Problems and sources of bad data

Sources Of Bad Data
(Percentage of survey respondents, multiple answers)

- Data entry by employees: 76%
- Data entry by customers: 25%
- Changes to source systems: 53%
- Data migration and conversion projects: 48%
- Mixed expectations by users: 46%
- External data: 34%
- Systems errors: 26%
- Other: 12%

Source: The Data Warehousing Institute

Problems Caused By Bad Data
(Percentage of survey respondents, multiple answers)

- Extra time to reconcile: 87%
- Delay deploying new systems: 64%
- Loss of credibility in system: 81%
- Lost revenue: 54%
- Extra costs (e.g. duplicate mail): 72%
- Customer dissatisfaction: 67%
- Compliance problems: 38%
- Other: 5%

Source: The Data Warehousing Institute
Data Integration and Migration

- System differences
- Poor information
- Incomplete Metadata
- Data Inconsistencies
- Rule Violations
- Data accessibility
- Time to analyse
- Focus on delivery not quality!!
Key to managing quality is understanding the process and data flows in your organization
“Middle managers spend an average of two hours per day looking for data.”

“The average IT manager spends 30% of their time trying to pin down information..”

“Data quality is broadening beyond its departmental roots into enterprise scope usage.”

— 2006 TDWI Data Quality Report finding
Why is Data Quality an Enterprise Issue?

• Data is a strategic asset
  – Driving Master Data Management initiatives
  – Driving Data Governance programs
• Significant operational cost savings
• Regulatory compliance
• Mergers and acquisitions
• Established enterprise technology initiatives
  – ERP
  – DW/BI
  – CRM
• Emerging enterprise technology initiatives
  – CDI
  – MDM
  – SOA
Mitigating the risk of poor data quality

- Stop making assumptions about your data
- Automate data analysis
- Discover hidden problems/information in your data
- Reduce data integration risks
- Ensure accurate and fit for purpose data
- Proactively measure, monitor and manage your data
Putting the Business Case Into Practice

**Action-Oriented Data Governance**

1. **Scope project**
2. **Define data metrics**
3. **Relate to business metrics**
4. **Capture baseline**
5. **Implement & track**
6. **Publish results**

Iterate

- **Key to demonstrable ROI**
- **Key to justifiable ROI**
3 Key Data Quality Steps

- **Data Profiling and Discovery**
  - **Automatic** discovery of potential data issues
  - Investigation of scale and causes
  - Determine scope and resolution of problems

- **Data Quality Cleansing**
  - Customize pre-built rules
  - Apply in batch during migrations
  - Real-time Integration into applications

- **Data Quality Monitoring**
  - Identify key data quality metrics
  - Monitor and measure over time
Why Data Profiling & Discovery?

Data Analysis and Design

Manual Profiling
50%
30%
20%

Automated Profiling
5%
30%
20%

90% Reduction (not including rework)

Project Cost/Time

Budget & Deadline
Data profiling to mitigate risk

- Automate data profiling
- Collaboration between information workers
  - Business analysts, data stewards, data analysts, decision makers
- Provide view of data issues in context
  - Technical data quality (structure, formats, integrity, content)
  - Business data quality (business rules, data domains, reference data, business context)
- Create standards and monitor
More information and Contact Details

White papers available at www.trilliumsoftware.com

Data Quality Essentials: For any Data-Intensive Project
Improving data migration with automated data profiling
The ROI of data quality
Building a tangible ROI for data quality
Realizing Master Data Management

Thank You

Email: Ed.wrazen@trilliumsoftware.com
Tel: UK (44) 118 940 7634