BEYOND BUSINESS PROCESS REENGINEERING (BPR):
DATA QUALITY ENGINEERING

Anton Raneses*
Defense Energy Support Center; anthony.raneses@dla.mil
For presentation at the Massachusetts Institute of Technology, Nov. 5-7, 2004

Executive Summary/Abstract: Process improvement tools such as Six Sigma could cost organizations like GE as much as 0.4% of its revenues. Consequently, organizations must constantly evaluate the cost-effectiveness of data quality tools. This presentation finds that beyond business process reengineering (BPR), data quality engineering in the form of failure mode and effects analysis (FMEA) could be utilized as a relatively inexpensive means to address data quality issues. The portrayed system integration example shows that on one hand, BPR provides a sound method of identifying data quality problems by way of gap analysis. On the other hand, FMEA allows user prioritization of such gaps as a function of risk, which likely leads to decreasing the bias in the calculation of the costs of systems change requests (SCRs) and problem trouble reports (PTRs) and providing a clearer picture of the bottom line.

*The views of the author do not necessarily represent those of the Department of Defense

Motivation #1
From a Total Data Quality Management (TDQM) standpoint, we are interested in capturing failure modes (data quality problems) in the software life cycle trend

Source: John Best, ETM 5291, Oklahoma State University
Motivation #2: Data quality problems in systems integration are, for the most part, captured via the SCR/PTR process, which is in serious need of improvement.

System Change Request (SCR) Form

<table>
<thead>
<tr>
<th>TO:</th>
<th>REQUESTOR</th>
<th>PRIORITY</th>
<th>REQUESTED DATE:</th>
<th>DSIO-SA (CM)</th>
<th>DSIO-SA (CM USE ONLY)</th>
<th>DATE RECEIVED</th>
<th>DATE CLOSED</th>
<th>PROBLEM AREA:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>05/05/2004</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>DED - Oracle Energy Downstream</td>
</tr>
<tr>
<td>TITLE: LOC DEV - Upload Daily OPIS Prices</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ASSOCIATED NUMBERS (7A4/DESD/issue/Scenario/Support Magic): DESC258</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

REQUESTOR:

<table>
<thead>
<tr>
<th>NAME (last, first):</th>
<th>Business Unit:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Thomas, M. / Zavros, T.</td>
<td>DESC-P</td>
</tr>
</tbody>
</table>

PHONe: COMMERCIAL: (703) 767-9378

We suggest that the data quality problems identified in our SCRs/PTRs are likely to be highly correlated with software quality risks.

List of Software Quality Risks

**Use cases:** working features fail  
**Robustness:** common errors are handled improperly  
**Performance:** slow system performance  
**Localization:** problems with time zones, currencies, etc.  
**Data integrity:** dbase becomes corrupted/accepts improper data  
**Usability:** software's interface is cumbersome or inexplicable  
**Volume/capacity:** the system fails at peak or sustained loads  
**Reliability:** at peak loads, the system crashes

We suggest that the data quality problems identified in our SCRs/PTRs are likely to be highly correlated with software quality risks.

*Source: Rex Black, 2002*
What is missing from this SCR/PTR table?

<table>
<thead>
<tr>
<th>Quality Risk</th>
<th>Failure Mode(s)</th>
<th>Priority</th>
</tr>
</thead>
<tbody>
<tr>
<td>Functionality</td>
<td>Can’t edit text.</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Can’t format text.</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Can’t handle tables.</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Can’t insert pictures.</td>
<td>3</td>
</tr>
<tr>
<td>Performance</td>
<td>Display more than two keystrokes behind.</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>File Ops longer than two seconds for large typical file.</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>File Ops longer than five seconds for large atypical file.</td>
<td>3</td>
</tr>
<tr>
<td>Compatibility</td>
<td>Can’t import Word files.</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Can’t import WordPerfect files.</td>
<td>3</td>
</tr>
</tbody>
</table>

Lacks notions of risk

**Severity:** How dangerous is a failure of the system stemming from this area?

**Priority:** How much does a failure of the system in this area compromise the value of the product to customers and users?

**Likelihood:** What are the odds that a user will encounter a failure in this area?

**Detection:** What are the odds that a failure in this area will escape detection?

*Source: Rex Black, 2002*

---

**Two Important Questions:**

- How important is the PTR/SCR (e.g., is the priority number realistic?)
- Not only should we consider how good the fix is, but also inquire “how much is the fix truly worth to us”? 

![Request Form](image)
Problem Statement – how could we improve the very mechanism (SCR/PTR) that addresses data quality problems in systems integration?

✓ BPR + gap analysis = a way/path to create and rank System Change Requests and Problem Trouble Reports (SCRs/PTRs)

✓ A method to rank SCRs/PTRs by their relative risk importance through data quality engineering (FMEA)

✓ Have the capability to obtain reliable pecuniary costs estimates of SCRs/PTRs

Objectives of this presentation

✓ Contribute to the literature on Information Quality (IQ)

✓ Showcase BPR & gap analysis efforts at a DoD agency

✓ Utilize the literature on FMEA and explain its usefulness to data quality

✓ Portray how might FMEA affect the bottom-line – the extent vendor-pricing estimates are likely biased
The Organization (DESC) – Executive Agent Candidate for Fuel Buys


- Consequently, a new automated fuel management system called FAS was obtained and implemented.

**Oil Energy Downstream (OED) + Oracle Government Financial (OGF) + Other = Fuel Automated System (FAS)**

- Systems Integration means transitioning from a Fortran-coded system to an ERP-based COTS, FAS.

TDQM literature emphasizes DMAIC Phases

- Define the business process
- Measure the performance of the core business process
- Analyze the process map to determine root causes of defects and means of improvement
- Improve by designing sustainable solutions to fix and prevent problems
- Control the improvements; keep new processes on check

Suggested Methodology: “Data Quality Engineering”/FMEA

FMEA is the study of failure mode and effects.

Failure Modes are sometimes described as categories of failure. A potential Failure Mode describes the way in which a product or process could fail to perform its desired function (design intent or performance requirements) as described by the needs, wants, and expectations of the internal and external Customers.

An Effect is an adverse consequence that the Customer might experience. The Customer could be the next operation, subsequent operations, or the end user.

Source: Haviland Consulting Group
### Severity Effect

<table>
<thead>
<tr>
<th>Probability of Failure</th>
<th>Detectability</th>
<th>Ranking</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hazardous without warning</td>
<td>Very High: Failure is almost inevitable</td>
<td>Absolute Uncertainty</td>
</tr>
<tr>
<td>Hazardous with warning</td>
<td>Medium High</td>
<td>Very Remote</td>
</tr>
<tr>
<td>Very High</td>
<td>High: Repeated failures</td>
<td>Remote</td>
</tr>
<tr>
<td>High</td>
<td>Low High</td>
<td>Very Low</td>
</tr>
<tr>
<td>Moderate</td>
<td>Moderate: Occasional failures</td>
<td>Low</td>
</tr>
<tr>
<td>Low</td>
<td>Low Moderate</td>
<td>Moderate</td>
</tr>
<tr>
<td>Very Low</td>
<td>High Low</td>
<td>Moderately High</td>
</tr>
<tr>
<td>Minor</td>
<td>Low: Relatively few failures</td>
<td>High</td>
</tr>
<tr>
<td>Very Minor</td>
<td>High Remote</td>
<td>Very High</td>
</tr>
<tr>
<td>None</td>
<td>Remote: Failure is unlikely</td>
<td>Almost Certain</td>
</tr>
</tbody>
</table>

### Probability of Failure

- **Hazardous without warning**
  - Very High: Failure is almost inevitable
- **Hazardous with warning**
  - Medium High
- **Very High**
  - High: Repeated failures
- **High**
  - Low High
- **Moderate**
  - Moderate: Occasional failures
- **Low**
  - Low Moderate
- **Very Low**
  - High Low
- **Minor**
  - Low: Relatively few failures
- **Very Minor**
  - High Remote
- **None**
  - Remote: Failure is unlikely

### Detectability

- Hazardous without warning: Absolute Uncertainty
- Hazardous with warning: Very Remote
- Very High: Remote
- High: Very Low
- Moderate: Low
- Low: Moderate
- Very Low: High
- Very High: Very High
- None: Almost Certain
From gap analysis, rank failure mode by priority
**Expected Cost Model**

Suppose that:
- $p_f$ = probability of a fault
- $p_d$ = probability that it escapes detection

Assuming that $p_f$ and $p_d$ are independent, the probability that the user receives the data quality problem or defect is $p_f \times p_d$.

- These probabilities can be estimated from the DESC Help Desk/ASG:
  - production records
  - customer records
  - sample inspection results

Source: John Best, ETM 5291, Oklahoma State University

---

**Expected Cost Model**

- Estimate cost per fault $C$

- Start with rough estimate based on
  - Internal scrap, rework
  - Warranty costs
  - Other Cost-Of-Poor-Quality factors
  - **Proxy for data quality failure (e.g. manual work-around)**

- If $n$ items are produced (yearly, monthly)

- Expected cost of data quality line items (PTRs/SCRs):
  $$EC = Cn \times p_f \times p_d$$

Source: John Best, ETM 5291, Oklahoma State University
## Estimating Occurrence

<table>
<thead>
<tr>
<th>Fault Occurrence</th>
<th>Data Quality 1</th>
<th>Data Quality 2</th>
<th>Data Quality 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probability</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5/10</td>
<td>0.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1/10</td>
<td>0.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5/100</td>
<td>0.05</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1/100</td>
<td>0.01</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>5/1000</td>
<td>0.005</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1/1000</td>
<td>0.001</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>5/10,000</td>
<td>0.0005</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>1/10,000</td>
<td>0.0001</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5/100,000</td>
<td>0.00005</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1/100,000</td>
<td>0.00001</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5/1,000,000</td>
<td>0.000005</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Source: John Best, ETM 5291, Oklahoma State University

## Estimating Undetection

<table>
<thead>
<tr>
<th>Undetection (escape)</th>
<th>Data Quality 1</th>
<th>Data Quality 2</th>
<th>Data Quality 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probability</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10/10</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8/10</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6/10</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3/10</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1/10</td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>5/100</td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>1/100</td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>5/1000</td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>1/1000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1/10,000</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Source: John Best, ETM 5291, Oklahoma State University
## Total Expected Cost Comparison

### Costs

<table>
<thead>
<tr>
<th>Failure Mode</th>
<th>Cost per Item</th>
<th>Month Volume</th>
<th>Probabilities</th>
<th>Expected Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>SCR1</td>
<td>50</td>
<td>20,000</td>
<td>0.01</td>
<td>0.1</td>
</tr>
<tr>
<td>SCR2</td>
<td>100</td>
<td>80,000</td>
<td>0.0005</td>
<td>0.05</td>
</tr>
<tr>
<td>SCR3</td>
<td>30</td>
<td>100,000</td>
<td>0.001</td>
<td>0.005</td>
</tr>
</tbody>
</table>

**Total Cost** $1,215

Source: John Best, ETM 5291, Oklahoma State University

## Results

### Data Quality Capture + Cost Effectiveness Tool

- **BPR** reveals data quality problems by way of gap analysis; leads to SCR/PTR documentation
- **FMEA** allows the user to incorporate risk in SCR/PTR costing

### Business Process Gap Analysis

<table>
<thead>
<tr>
<th>Business Process</th>
<th>Gap Analysis</th>
<th>Nominal Cost ($)</th>
<th>Estimated Cost ($)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ordering</td>
<td>Suspension SCR</td>
<td>biased/unbiased</td>
<td>closer to true costs</td>
</tr>
<tr>
<td>Receipting</td>
<td>Payment problems</td>
<td>biased/unbiased</td>
<td>closer to true costs</td>
</tr>
<tr>
<td>Pricing</td>
<td>Failed escalator</td>
<td>biased/unbiased</td>
<td>closer to true costs</td>
</tr>
<tr>
<td>Taxes</td>
<td>Tax change</td>
<td>biased/unbiased</td>
<td>closer to true costs</td>
</tr>
</tbody>
</table>
Next Steps

BPR & gap analysis were large steps for DESC, but it is just the beginning...

- Gather and compile the data (e.g. collect \( n \) items that are produced yearly, monthly, etc.) for the purposes of Data Quality Capture
- Utilize Application System Group (ASG)/Help Desk Ticket info
  - Direct Delivery and Bulk Fuel transactions
  - Inventory/Stock Control statistics
  - Other operational work
- Integrate Data Quality Engineering (FMEA) in Gap Analysis
  - Conduct frequency analysis
  - Use sensitivity analysis

Further Work

Summary

Problem Statement
How could we improve the very mechanism (SCR/PTR) that addresses data quality problems in systems integration?

Methodology
BPR + Gap Analysis = Part of the solution
Data Quality Engineering (FMEA) = Proposed solution

Results
Incorporates notion of risk in PTR/SCR costs
Likely decreases biased estimates

Further work/Challenges at DESC
This is merely a process presentation. The challenge ahead is in the execution of the proposed methodology.
References