Acxiom’s Data Quality Scorecard Solution

MIT IQ Industry Symposium
Cambridge, Massachusetts, USA
July 18-19, 2007
You Cannot Improve What You Cannot Measure

“Without a systematic way to conceptualize and address the data quality issue, organizations are left to grapple with this problem in an ad hoc, piece-meal manner.”

– MIT Total Data Quality Management Program
Overview

Acxiom’s Data Quality Scorecard Solution (DQSS) helps companies determine if their data is incomplete, inaccurate or invalid. Companies equipped with this kind of knowledge can make better data decisions – such as whether to decrease the amount paid to vendors, augment their data or eliminate redundant data processing.
Purpose

- Acxiom’s Data Quality Scorecard Solution helps clients understand the condition of their customer, prospect or operational data.
- Whether high or low, data quality is always relative to the data’s intended use. We approach data quality with a strategic, multi-dimensional view appropriate to the data’s business use. This strategic management also minimizes “accidental” discovery of deficiencies.
# Data Quality

<table>
<thead>
<tr>
<th>Data Quality Challenges</th>
<th>Data Quality Management Benefits</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data quality issues often recognized too late in the data integration process</td>
<td>DQ issues exposed through <strong>monitoring</strong> rather than accidental discovery</td>
</tr>
<tr>
<td>Complex, dynamic data environments introduce many points of failure</td>
<td>Unnecessary production costs are reduced through automation</td>
</tr>
<tr>
<td>Business impact of data is often not reflected in “off-the-shelf” solutions</td>
<td>“Fitness for use” is built into the daily DQ management. Key metrics are monitored through a central repository</td>
</tr>
<tr>
<td>Data sources introduce changes that impact data stores or individual marts</td>
<td>Changes are exposed early, reducing the impact on your data marts. Periodic, quantifiable measurements ultimately reduce expense for poor inbound data quality</td>
</tr>
<tr>
<td>Tactical solutions solve immediate problems without identifying and fixing underlying issues</td>
<td>DQ infrastructure supplements “one-time” assessments and pinpoints where the problems originate</td>
</tr>
<tr>
<td>Regulatory non-compliance or decreasing customer confidence is often rooted in vague data quality definitions</td>
<td>DQ management system provides statistical evidence</td>
</tr>
</tbody>
</table>
Solution Components

• Professional consulting with Acxiom data quality specialists to research existing data quality processes, design client-specific data quality metrics and create a Data Quality Scorecard that meets their business needs
• A complete repository system for input, storage, scoring, archiving and dashboard-reporting of data quality metrics and scorecards
• Optional data quality products (Acxiom’s Data Quality Toolkit) that can be automatically integrated into the client’s scorecard system
Strategic Approach: Total Data Quality Management (TDQM)
DQSS and TDQM Working Together

Professional Services

• Data Quality Assessments and Consulting
• Metrics Defined for:
  – Data Access
  – Data Accuracy
  – Data Completeness
  – Data Consistency
  – Grouping Accuracy

Measurement Utilities

• Data Validator
• NameCheck
• AccuCheck
• Opticx®
• AddressAbility®
• Hygiene
• Other Custom Applications

Improvement Potential

• Data Augmentation
• Analytic Capabilities
• Data Integration
• Etc.

Professional Services

• Data Quality Consulting
  – Analyzing Failures
  – Utilizing Advantages
• Metric Trending
DQSS in Practice (Single Touch Point or Database)

**Discovery, Analysis and Metric Design Services**
- Define Data Quality Dimensions
- Define Metrics, Goals, Thresholds

**Acxiom CII Services: Batch or Grid-Enabled**

- **Data Accuracy**
  - Acxiom’s AccuCheck Benchmark Comparison Process
- **Data Access**
  - Throughput Metrics
- **Data Completeness**
  - Acxiom’s Data Validator
- **Data Consistency**
  - Acxiom’s Data Validator (with Trending or Comparator Plug-in)
- **Grouping Accuracy**
  - Matching Logic / Business Rule Process

**Data Quality Scorecard Repository System**
- Scorecard Reports generated and displayed
- Measurement values and quality scores calculated and stored
- Data Quality trending graphs generated
- Key business metrics monitored regularly

Source Data

DQSR

Metric Measurements

Copyright © 2007 Acxiom Corporation. All Rights Reserved.
DQSS Across the “Data Factory” (Multiple Enterprise Touch Points)

Data Flow(s)

Touch Point Metrics
- Consistency
- Access
- Completeness
- Accuracy
- Grouping

Measurement Tool(s)

DQSR System

Touch Point Metrics
- Consistency
- Access
- Completeness
- Accuracy
- Grouping

Measurement Tool(s)

Project(s)

IMPROVEMENT

ANALYSIS

Scorecards
DQSS Across the “Data Factory”

• **Business Need:** Improve the approach to data quality assurance on the client’s credit card services, “Operational Data Store.”

• **Problem:** Current data management lacked methods to measure data quality in a consistent manner so that issues could be addressed proactively. This created situations resulting in:
  - lost revenue opportunities
  - unnecessary source data expense
  - decreased customer satisfaction
  - lack of end-user confidence in the system of record data
  - additional regulatory scrutiny and fines
DQSS Across the “Data Factory”

- **Project Strategy:** Apply an industry best practice approach in the form of a **Data Quality Scorecard Solution** to measure, analyze, and ultimately improve data quality.

- **Objective:** Identify and implement custom data quality metrics for the client’s operational data infrastructure and use metrics to identify improvement opportunities.

- **Scope:** Project consisted of consulting & analysis services and measurement system including Scorecard Repository.
Data Quality Scorecard Solution (DQSS)

Discovery, Analysis and Metric Design Services
• Define Data Quality Dimensions
• Define Metrics, Goals, Thresholds

Define

Measure

Source Data

Data Accuracy

Data Access

Data Completeness

Data Consistency

Grouping Accuracy

Output Data

Statistics

ANALYZE & MANAGE

DQSR

Scorecards

DQSR Web System
• Scorecard Reports generated and displayed
• Measurement values and quality scores calculated and reported
• Data Quality reports and trending graphs generated
• Key business metrics monitored regularly

DEFINE

MEASURE

ANALYZE & MANAGE

Copyright © 2007 Acxiom Corporation. All Rights Reserved.
Objective: Acxiom DQ consultants gathered information about the selected data, processes, and client data quality goals for Phase 1.

Methodology: Conducted interviews with the client’s DQ Team; determined the data quality goals and requirements; prioritized the metrics with rankings. For the highest-ranking issues, Acxiom recommended measurement methods.

Client Discovery Participation:
• Identified the principal DQ Scorecard Champion who would have the authority to approve (sign-off on) the final set of metrics
• Identified members of their DQ team with knowledge to assist in the analysis (e.g., data stewards, SMEs, or business/end-users)
• Prepared discussion around existing or desired data quality requirements and goals.
• Educated their DQ Team using Acxiom-provided reference material in advance of the on-site consulting visit.

Objective: Reached agreement (sign-off) on the metric definitions for the projects and agreement on which of those would be implemented in the production process using Acxiom Products & Solutions.

Methodology: Consultants prepared the set of metrics and submitted to client’s DQ Scorecard Champion for approval.

Deliverable: Metric Design Document and Metric Calculator Algorithms were produced for 22 metrics spanning 15 database tables.
**Define**

**Measure**

**Analyze & Manage**

---

**Metric Design Document**

**System:** OIS  
**Touchpoint:** OIS Core Tables  
**Metric Interval (Frequency of calculation):** Daily

**Metric Description:** Verification that the current account balance in the Account table is equal to the current balance in the Transaction_Balance table.

**Goal:** 0% (variance)  
Max: 0%  
Failure: 10%

**Table(s):** Account  
Transaction_Balance

**Column(s):** Account.Current_Balance  
Transaction_Balance.Current_Balance

**Algorithm Requirements:**

\[
\text{Sum of Account.Current_Balance where Current_Balance > 0} = \text{Sum of Transaction_Balance.Current_Balance. Metric Value is the difference in sums expressed as a percentage of the Account.Current_Balance.}
\]

**Technical Specification (Metric Calculator Design):**

Two Comparator report values:  
\[X = \text{Value of SUM(Account.Current_Balance) where Current_Balance > 0}\]  
\[Y = \text{Value of SUM(Transaction_Balance.Current_Balance) where Account.Key > 0}\]  
\[M8 = ((\text{ABS}(x - y)) / x) * 100\]
## Metric Classification

Metrics were prioritized and ranked, then classified according to DQ Dimension and recommended measurement tool.

<table>
<thead>
<tr>
<th>Metric ID</th>
<th>Metric Name</th>
<th>DQ Dimension</th>
<th>Measurement Tool</th>
</tr>
</thead>
<tbody>
<tr>
<td>M8</td>
<td>Account Balance Integrity</td>
<td>Accuracy</td>
<td>DV-Comparator</td>
</tr>
<tr>
<td>M11</td>
<td>Cardholder Demographic Verification, Postal Area Code</td>
<td>Consistency</td>
<td>Data Validator</td>
</tr>
<tr>
<td>M25</td>
<td>Event Distribution</td>
<td>Consistency</td>
<td>DV-Trending</td>
</tr>
</tbody>
</table>
DQSS Implementation

Logical Context Diagram

- Data Validator
- Trending
- Comparator

Report Library Structure
- Tool Folder
- Table Folder
- Reports

Start Script

Application Layer

Web Access

Report Generator

Score Calculator

Access Layer

Metric Repository

Scores
Metric
TP
User
Data
Data
Log
Data Quality Scorecard Repository
Executive Summary Dashboard
“M8: Account Balance Integrity”

Accuracy: Verification that the current balance in the Account table is equal to the current balance in the Transaction_Balance table.
# M8 Data Validator-Comparator Report

**Total Records:** 7 of 7

**Condition(s):**
- CURRENT_BALANCE > 0
- ACCT_FLAG > 1

**Transaction Balance, Current Balance**

<table>
<thead>
<tr>
<th>Description</th>
<th>Summary</th>
</tr>
</thead>
<tbody>
<tr>
<td>Transaction Balance</td>
<td>84146</td>
</tr>
<tr>
<td></td>
<td>103201</td>
</tr>
<tr>
<td></td>
<td>75625</td>
</tr>
<tr>
<td></td>
<td>46250</td>
</tr>
<tr>
<td></td>
<td>75613</td>
</tr>
<tr>
<td></td>
<td>124023</td>
</tr>
</tbody>
</table>

**Account Current Balance**

<table>
<thead>
<tr>
<th>Description</th>
<th>Summary</th>
</tr>
</thead>
<tbody>
<tr>
<td>Account Current Balance</td>
<td>84146</td>
</tr>
<tr>
<td></td>
<td>103201</td>
</tr>
<tr>
<td></td>
<td>75605</td>
</tr>
<tr>
<td></td>
<td>46250</td>
</tr>
<tr>
<td></td>
<td>75613</td>
</tr>
<tr>
<td></td>
<td>164023</td>
</tr>
</tbody>
</table>

**Job Summary**

**Description** | **Summary**
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Version</td>
<td>2.1</td>
</tr>
<tr>
<td>Date/Time</td>
<td>2006-05-07 14:14:03</td>
</tr>
</tbody>
</table>

**Description** | **Summary**
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Version</td>
<td>2.1</td>
</tr>
<tr>
<td>Date/Time</td>
<td>2006-05-07 14:13:56</td>
</tr>
</tbody>
</table>

---

*Copyright © 2004 Acxiom Corporation*  
*All rights reserved*  
*Data Validator Report*

---

*Proceedings of the MIT 2007 Information Quality Industry Symposium*
M8 Metric as entered into DQSR

```
MetricID: 194
System: Financial Services
Touchpoint: Fraud Management
Metric Name: M8
Dimension: Data Accuracy
Goal: 0
Failure: 10
Max: 0
If LessThan: ✓
Interval: Daily
Include in GPA: ✓
Metric Description: Account Balance Integrity. Verification that the current balance in the Account table is equal to the current balance in the Transaction_Balance table. Sum of
```

“M8” Metric Scorecard displayed in DQSR
M11: Demographic Verification/Postal Area Code

Completeness: Every cardholder has a valid postal area code
### “M11” Data Validator Reports

Here is a snippet of the report:

```
Time: Tue May 09 10:20:23 2006
Program Version: 2.0
File: stdin
Record Length: 15

Total Number of Records: 1000

<table>
<thead>
<tr>
<th>Element Name</th>
<th>Record Count</th>
<th>% of Total</th>
<th>Value Found</th>
</tr>
</thead>
<tbody>
<tr>
<td>POSTAL_AREA_CD</td>
<td>502</td>
<td>50.20%</td>
<td>Blank</td>
</tr>
<tr>
<td></td>
<td>39</td>
<td>3.90%</td>
<td>Invalid</td>
</tr>
<tr>
<td></td>
<td>495</td>
<td>45.90%</td>
<td>Valid ()</td>
</tr>
</tbody>
</table>
```
"M11" Data Validator Reports

<table>
<thead>
<tr>
<th>Element Name</th>
<th>Error Count</th>
<th>% of Total</th>
<th>Element Content</th>
</tr>
</thead>
<tbody>
<tr>
<td>POSTAL_AREA_CD</td>
<td>3</td>
<td>0.20%</td>
<td>7221</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0.10%</td>
<td>857964257</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0.10%</td>
<td>72205-2687</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0.10%</td>
<td>75205-3</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0.10%</td>
<td>72212-2227</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0.10%</td>
<td>75024-1</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0.10%</td>
<td>85020-138</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0.10%</td>
<td>752240</td>
</tr>
</tbody>
</table>
### “M11” Data Validator Reports

```
Time: Tue May 09 10:20:23 2006
Program Version: 2.0
File: stdin
Record Length: 15

<table>
<thead>
<tr>
<th>Record Number</th>
<th>Field ID</th>
<th>Element Name</th>
<th>Error Count</th>
<th>Content</th>
</tr>
</thead>
<tbody>
<tr>
<td>12</td>
<td>001</td>
<td>POSTAL_AREA_CD</td>
<td>1</td>
<td>7221</td>
</tr>
<tr>
<td>14</td>
<td>001</td>
<td>POSTAL_AREA_CD</td>
<td>2</td>
<td>7221</td>
</tr>
<tr>
<td>15</td>
<td>001</td>
<td>POSTAL_AREA_CD</td>
<td>3</td>
<td>857964257</td>
</tr>
<tr>
<td>19</td>
<td>001</td>
<td>POSTAL_AREA_CD</td>
<td>4</td>
<td>72205-2687</td>
</tr>
<tr>
<td>20</td>
<td>001</td>
<td>POSTAL_AREA_CD</td>
<td>5</td>
<td>75205-</td>
</tr>
<tr>
<td>22</td>
<td>001</td>
<td>POSTAL_AREA_CD</td>
<td>6</td>
<td>72212-2227</td>
</tr>
<tr>
<td>43</td>
<td>001</td>
<td>POSTAL_AREA_CD</td>
<td>7</td>
<td>72212-2227</td>
</tr>
<tr>
<td>50</td>
<td>001</td>
<td>POSTAL_AREA_CD</td>
<td>8</td>
<td>75024-</td>
</tr>
<tr>
<td>53</td>
<td>001</td>
<td>POSTAL_AREA_CD</td>
<td>9</td>
<td>75224-</td>
</tr>
<tr>
<td>68</td>
<td>001</td>
<td>POSTAL_AREA_CD</td>
<td>10</td>
<td>72212-2227</td>
</tr>
<tr>
<td>76</td>
<td>001</td>
<td>POSTAL_AREA_CD</td>
<td>11</td>
<td>7221</td>
</tr>
<tr>
<td>80</td>
<td>001</td>
<td>POSTAL_AREA_CD</td>
<td>12</td>
<td>85020-1328</td>
</tr>
<tr>
<td>83</td>
<td>001</td>
<td>POSTAL_AREA_CD</td>
<td>13</td>
<td>752240</td>
</tr>
</tbody>
</table>
```
“M11” Metric as entered into DQSR

Add, Edit, Delete, Or View Metrics

- MetricID: 209
- System: Financial Services
- Touchpoint: Marketing
- Metric Name: M11
- Dimension: Data Consistency
- Goal: 90
- Failure: 50
- Max: 100
- If Less Than: 0
- Interval: Daily
- Include in GPA: 
- Metric Description: Cardholder Demographic Verification, Postal Area Code
“M11” Metric Scorecard displayed in DQSR
M25: Event Distribution

**Consistency:** Verification that all “Event Types” fall within the expected and consistent distribution of valid values.

---

Copyright © 2007 Acxiom Corporation. All Rights Reserved.
"M25" Trending Report

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Blank</td>
<td>0</td>
<td>0.00%</td>
<td>0</td>
<td>0.00%</td>
<td>0</td>
<td>0.00%</td>
</tr>
<tr>
<td>0</td>
<td>230</td>
<td>22.01%</td>
<td>200</td>
<td>19.90%</td>
<td>30</td>
<td>15.00%</td>
</tr>
<tr>
<td>1</td>
<td>203</td>
<td>19.43%</td>
<td>203</td>
<td>20.20%</td>
<td>0</td>
<td>0.00%</td>
</tr>
<tr>
<td>2</td>
<td>201</td>
<td>19.23%</td>
<td>201</td>
<td>20.00%</td>
<td>0</td>
<td>0.00%</td>
</tr>
<tr>
<td>3</td>
<td>200</td>
<td>19.14%</td>
<td>200</td>
<td>19.90%</td>
<td>0</td>
<td>0.00%</td>
</tr>
<tr>
<td>4</td>
<td>211</td>
<td>20.19%</td>
<td>201</td>
<td>20.00%</td>
<td>10</td>
<td>4.98%</td>
</tr>
</tbody>
</table>
“M25” Metric as entered into DQSR
“M25” Metric Scorecard displayed in DQSR

DQSR SYSTEM

- DQSR System Home
- Log Off DQSR System

REPORT OPTIONS
- Reports Dashboard
- Metric Calculator Logs

SCORECARD OPTIONS
- View Scorecards
  - Metric Scorecards
  - Touchpoint Scorecards
- Scorecard Administration

ADMINISTRATIVE OPTIONS
- Repository Administration

SUPPORT INFORMATION
- DQSR Reference Guide
- About DQSR System
- Change Your Password

Copyright © 2005 Axiom Corporation
Axiom’s Privacy Policy

PG 477

Copyright © 2007 Axiom Corporation. All Rights Reserved.