Comprehensive Data Quality Improvement through Data Governance
A Management Overview

Introducing MIKE2.0
An Open Source Methodology for Information Development
http://www.openmethodology.org
Data Quality Improvement through Data Governance

**Agenda**

- Defining Data Quality and Data Governance
- Implications of poor Data Governance
- Implementing a Successful Data Governance Programme
  - Challenges
  - Guiding Principles
  - Delivery Methodology
- Lessons Learned
- Discussion Questions Related to Data Governance
Data Quality Improvement
Scope Within BearingPoint's IM Solution Suite

Information Management Solution Suite

Delivered through a Collaborative Approach with the IM Profession and our Alliance Vendors

Enterprise Information Management

Supported by Solution Capabilities that provide a foundation for Suite Delivery

- BI and EPM
- Information Asset Management
- Access, Search and Content Delivery

Business Solutions

- Enterprise Data Management
- Enterprise Content Management

Commercial & Open Source Product Solutions

Information Strategy, Architecture and Governance

Sets the new standard for Information Development through an Open Source Offering

© 2008 BearingPoint, Inc.
Data Quality Improvement
How we Define Data Quality

Data Quality Management is a complex topic that involves more than just the accuracy of data. Data Quality is typically measured across quantitative dimensions and qualitative dimensions.

### Measurable Dimensions for Data Quality

<table>
<thead>
<tr>
<th>Dimension</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
<td>Does the data accurately represent reality or a verifiable source?</td>
</tr>
<tr>
<td>Integrity</td>
<td>Do broken links exist between data that should be related?</td>
</tr>
<tr>
<td>Consistency</td>
<td>Is there a single representation of data?</td>
</tr>
<tr>
<td>Completeness</td>
<td>Is any key information missing?</td>
</tr>
<tr>
<td>Uniqueness</td>
<td>Is the data value unique? i.e. no duplicate values or records</td>
</tr>
<tr>
<td>Accessibility</td>
<td>Is the data easily accessible, understandable, and used consistently?</td>
</tr>
<tr>
<td>Precision</td>
<td>Is data stored with the precision required by the business?</td>
</tr>
<tr>
<td>Timeliness</td>
<td>Is the information update frequency adequate to meet the business requirements?</td>
</tr>
</tbody>
</table>

### Intangible Dimensions

<table>
<thead>
<tr>
<th>Dimension</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Relevance</td>
<td>Every piece of information stored is important in order to get a true business representation</td>
</tr>
<tr>
<td>Usability</td>
<td>The stored information is usable by the organization with ease</td>
</tr>
<tr>
<td>Usefulness</td>
<td>The stored information is applicable for the organization</td>
</tr>
<tr>
<td>Believability</td>
<td>The level to which the data is regarded as true and credible</td>
</tr>
<tr>
<td>Unambiguous</td>
<td>Each piece of data has a unique meaning, and can be easily comprehended</td>
</tr>
<tr>
<td>Objectivity</td>
<td>Data is objective, unbiased and impartial i.e., it does not depend on the judgment, interpretation, or evaluation of people</td>
</tr>
</tbody>
</table>

Data Quality Management involves more than just addressing historical data quality issues through data profiling and re-engineering. It involves preventing these issues from occurring in the first place.
Data Quality Improvement
Data Governance is the Key to Data Quality Improvement

There are varying definitions of the term "Data Governance". Whilst there is consensus that Data Governance includes Data Quality Management, it is difficult to get a consistent definition even at a high level. There are 3 primary reasons for this:

- **Data Quality is a composite topic** that can be measured across several dimensions

- **Preventing Data Quality issues is complex**, sometimes involving changes to source systems, business processes and supporting technology. Issues for some systems or users may not present a problem for others

- **We define Data Governance to mean more than just Data Quality.** It includes standards, policies, processes, delivery best practices, organizational efficiency and agility

Our position is that complex Data Quality issues cannot be addressed without a comprehensive Data Governance programme.
Data Quality Improvement
Data Governance is the Key to Data Quality Improvement

What makes up a Comprehensive Data Governance Programme?

- Overall vision for Improvement
- Programme implementation plan
- Linkage of business to technology and strategic to tactical
- Flexibility to change
- Open and Common standards
- Strategic Conceptual links through to incremental Solution Architecture
- End-to-End Data Security
- Information Development Organization
- Data is an embedded competency in all staff
- Data Stewards and named Senior Stakeholders across the organization
- Metrics definition
- Data Standards
- Policy framework
- Continuous improvement of policies over time
- Properly defined controls
- Correction processes
- Root cause analysis
- Best practices and methods
- Focus on improvement by starting with Key Data Elements
- A Supply Chain based approach to Data Governance
The Impact of Poor Data Governance
What are the Impacts to Project Delivery?

Data Quality is and has been a primary problem in project failures – and the issue isn’t going away

- According to the Standish Group, in 1998, 74 percent of all data migration projects either overran or failed, resulting in almost $100 billion in unexpected costs.
- In a survey of 300 IT executives conducted by Information Week, the majority of the respondents (81 percent) said, "improving (data) quality was their most important post-year 2000 technology priority".
- Data Quality issues lead to 87% of projects requiring extra time to reconcile data – TDWI Data Quality Survey, December, 2001.
- Data Quality issues lead to lost credibility within a system in 81% of cases – TDWI Data Quality Survey, December, 2001.
- According recent studies (2005) to the Gartner and Meta Groups, 55 – 70% of CRM and 70% of Data Warehouse project failures are due to data quality issues.
- Through 2005, more than 50% of CRM deployments will suffer limited acceptance, if not outright failure, because of the lack of attention to data quality issues., Gartner, 2005.
- In a recent report, Gartner predicted 50% of data warehouse projects through 2007 will have limited acceptance or be outright failures, as a result of lack of attention to data quality issues.

Despite the Impact of Poor Data Quality, Most Organizations Do Not Address it Comprehensively

What's the status of your organization's data governance initiative?

<table>
<thead>
<tr>
<th>Status</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>No plans</td>
<td>42%</td>
</tr>
<tr>
<td>Under Consideration</td>
<td>33%</td>
</tr>
<tr>
<td>Design or implementation phase</td>
<td>17%</td>
</tr>
<tr>
<td>Deployed</td>
<td>8%</td>
</tr>
</tbody>
</table>


As per the new UK Data Protection Act, "information must be accurate, up-to-date and held for no longer than necessary". Under this new law, companies have to comply with rules for manual and electronic processing of personal data, such as name, addresses and dates of births.
The Impact of Poor Data Governance

Data Governance Horror Stories

If you have a poorly implemented approach to Data Governance, you too can make the news

Poor error check checking and lack of safeguards can result in big loss

- An employee of Mizuho Securities Co. Ltd typed by accident an order to sell 610,000 shares at 1 yen, instead of an order to sell 1 share at 610,000 yen. The Tokyo Stock Exchange initially blocked attempts to cancel the order, resulting in a **net loss of 347 million US dollars** to be shared between the exchange and Mizuho – “Costly Error by Japanese Firm,” The New York Times, December 9, 2005 by Reuters

- On October 2, 2002, Bear, Stearns placed an order to sell $4 billion in stocks – but the order was intended to be for only $4 million. While the orders were cancelled within minutes, **$622 million in stock had been sold in error** and Bear, Stearns needed to pay the difference in price to buy it back. – “Erroneous Order for Big Sales Briefly Stirs Up the Big Board,” New York Times, October 3, 2002 by Floyd Norris

Resilience to high volume trading

- The Tokyo stock exchange was not able to cope with high level of trading volume in 01/11/2005 and had to be **suspended for four and a half hours**

Data Security Issues

- Citigroup has incurred two separate losses of magnetic tapes with customer names, addresses, account numbers and balances. A tape with information about 120,000 Japanese customers was lost en route to Singapore and tapes with personal information of almost 4 million American customers was lost in the US. – “The Scramble to Protect Personal Data,” The New York Times, June 9, 2005 by Tom Zeller, Jr.

Inconsistencies between Operational and Reporting Systems

- Statistics South Africa, the government agency that reports official economic data, admitted last June that it had greatly overstated inflation for several months. The result: interest rates stayed too high for too long. As a result, some foreign investors may have been deterred by the low reported rates. – “Oops, I did it again,” The Economist, June 17, 2004
Improving Data Governance
What are the Benefits to Improving Data Governance

**Most Organisations are Currently here**

<table>
<thead>
<tr>
<th>Key Issues</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Inconsistencies Exist Across Systems</td>
</tr>
<tr>
<td>Staff Time Wasted Massaging Data</td>
</tr>
<tr>
<td>Fragmented View of Customer Exists</td>
</tr>
<tr>
<td>Customer Records are Duplicated</td>
</tr>
<tr>
<td>Accuracy Issues Exist for Key Data Elements</td>
</tr>
<tr>
<td>Inefficient Use of Resources</td>
</tr>
<tr>
<td>11th Hour Delivery Risks Due to Bad Data</td>
</tr>
</tbody>
</table>

**Improved Data Governance**

**Your Goal is to get to here**

<table>
<thead>
<tr>
<th>Desired State</th>
</tr>
</thead>
<tbody>
<tr>
<td>Key Data Elements in Synch Across Systems</td>
</tr>
<tr>
<td>MIS Staff Spends Time Analyzing not Verifying</td>
</tr>
<tr>
<td>Integrated View of Customer Across the Bank</td>
</tr>
<tr>
<td>Customer Records are Unique</td>
</tr>
<tr>
<td>Required Accuracy Levels are Consistently Achieved</td>
</tr>
<tr>
<td>Improved Operational Efficiency</td>
</tr>
<tr>
<td>Quantitative Knowledge of Quality Issues</td>
</tr>
</tbody>
</table>

**Impetus to Change**

- Drive Down Costs in the Business
- Improve Competitive Position
- Meet Regulatory Requirements

© 2008 BearingPoint, Inc.
Data Quality Improvement
Why is it so Hard?

Dispute the tremendous cost of Data Quality issues, most organisations are struggling to addresses their Data Quality issues. We believe there are 5 primary reasons why they are failing:

- **Our Systems are more complex than ever before.** Companies now have more information and are conducting more integration between systems than ever before. New regulations, M&As, globalization and increasing customer demands mean that information management challenges are increasingly formidable.

- **Silo-ed, short-term project delivery focus.** As projects are often funded at a departmental level, they don’t account for the impacts of how data will be used by others. Data flows between systems and the design of these connection points must go across strict project boundaries.

- **Traditional development methods do not give enough focus to data management.** Many projects are focused more on function and feature than on information – the desire to build new functionality has resulted in a information being left by the wayside.

- **Data Quality issues are Hidden and Persistent.** Data quality issues can exist unnoticed for some time, although some users may suspect the data in the systems they rely on to make their decisions is not accurate, complete, current, valid, or consistent. This data can then get propagated to other systems as we increase connectivity. Organizations tend to underestimate the data quality issues in their systems.

- **Data Quality is Fit for Purpose.** It is difficult for users of downstream systems to improve the data quality of their system because the data they derive information from is entered via customer facing operational systems. These customer facing system operators do not have the same incentive to maintain high data quality and they are focused on entering the data quickly and without rejection by the system at the point of entry. **It is often when data is integrated, summarized, standardized and used in another context that quality issues begin to surface.**

A comprehensive Data Governance programme must be defined to meet these challenges.
The MIKE2.0 Methodology
An Open Source Methodology for Information Development

What is MIKE2.0?

- MIKE stands for Method for an Integrated Knowledge Environment
- MIKE2.0 is our comprehensive methodology for Enterprise Information Management
- MIKE2.0 brings together important concepts around Open Source and Web 2.0
- The open source version of MIKE2.0 is available at: http://www.openmethodology.org

Key Constructs within MIKE2.0

- SAFE (Strategic Architecture for the Federated Enterprise) is the architecture framework for the MIKE2.0 Methodology
- Information Development is the key conceptual construct for MIKE2.0 – develop your information just like applications

MIKE2.0 provides a Comprehensive, Modern Approach

- Scope covers Enterprise Information Management, but goes into detail in areas to be used for more tactical projects
- Architecturally-driven approach that starts at the strategic conceptual level, goes to solution architecture
- A comprehensive approach to Data Governance, Architecture and strategic Information Management

MIKE2.0 provides a Collaborative, Open Source Methodology for Information Development

- Balances adding new content with release stability through a method that is easier to navigate and understand
- Allows non-BearingPoint users to contribute
- Links into BearingPoint's existing project assets on our internal knowledge management systems
- Unique approach, we would like to make this "the standard" in the new area of Information Development
The MIKE2.0 Methodology
The 5 Phases of MIKE2.0

Information Development through the 5 Phases of MIKE2.0

Strategic Programme
Blueprint is done once

Phase 1
Business Assessment

Phase 2
Technology Assessment

Continuous Implementation Phases

Increment 1
Increment 2
Increment 3

Phase 3, 4, 5

Begin Next Increment

Improved Governance and Operating Model

© 2008 BearingPoint, Inc.
The MIKE2.0 approach for improving Data Governance goes across all 5 phases of the methodology. The most critical activities for improving Data Governance are as follows:

- Activity 1.4 Organizational QuickScan
- Activity 1.6 Data Governance Sponsorship and Scope
- Activity 1.7 Initial Data Governance Organization
- Activity 2.7 Data Governance Policies
- Activity 2.8 Data Standards
- Activity 3.5 Business Scope for Improved Data Governance
- Activity 3.6 Enterprise Information Architecture
- Activity 3.7 Root Cause Analysis on Data Governance Issues
- Activity 3.8 Data Governance Metrics
- Activity 3.11 Data Profiling
- Activity 3.12 Data Re-Engineering
- Activity 5.11 Continuous Improvement – Compliance Auditing
- Activity 5.12 Continuous Improvement – Standards, Policies and Processes
- Activity 5.13 Continuous Improvement – Data Quality
- Activity 5.14 Continuous Improvement – Infrastructure
- Activity 5.15 Continuous Improvement – Information Development Organization
- Activity 5.16 Continuous Improvement – MIKE2.0 Methodology

Other MIKE2.0 Activities are also relevant, but these are particularly focused on Data Governance.
The MIKE2.0 Methodology
The First Step: Assessing Data Governance Levels

Information Development through the 5 Phases of MIKE2.0

Strategic Programme Blueprint is done once

Phase 1 Business Assessment
Phase 2 Technology Assessment

Continuous Implementation Phases

Increment 1
Increment 2
Increment 3

Activity 1.4 Organizational QuickScan for Information Development

<table>
<thead>
<tr>
<th>Activity</th>
<th>Responsible</th>
<th>Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.4.1 Assess Current-State Application Portfolio</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.4.2 Assess Information Maturity</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.4.3 Assess Economic Value of Information</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.4.4 Assess Infrastructure Maturity</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.4.5 Assess Key Current-State Information Processes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.4.6 Define Current-State Conceptual Architecture</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.4.7 Assess Current-State People Skills</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.4.8 Assess Current-State Organizational Structure</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.4.9 Assemble Findings on People, Organization and its Capabilities</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Improved Governance and Operating Model

© 2008 BearingPoint, Inc.
The MIKE2.0 Methodology
The First Step: Assessing Data Governance Levels

Activity 1.4 Organizational QuickScan for Information Development

<table>
<thead>
<tr>
<th>Activity</th>
<th>Responsible</th>
<th>Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.4.1 Assess Current-State Application Portfolio</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.4.2 Assess Information Maturity</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.4.3 Assess Economic Value of Information</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.4.4 Assess Infrastructure Maturity</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.4.5 Assess Key Current-State Information Processes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.4.6 Define Current-State Conceptual Architecture</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.4.7 Assess Current-State People Skills</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.4.8 Assess Current-State Organizational Structure</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.4.9 Assemble Findings on People, Organization and its Capabilities</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Phase 1 – Business Assessment and Strategy Definition Blueprint

1.1 Strategic Mobilisation
1.2 Enterprise Information Management Awareness
1.3 Overall Business Strategy for Information Development
1.4 Organisational QuickScan for Information Development
1.5 Future State Vision for Information Management
1.6 Data Governance Sponsorship and Scope
1.7 Initial Data Governance Organisation
1.8 Business Blueprint Completion
1.9 Programme Review
The MIKE2.0 Methodology
Data Profiling: Quantitatively Understand your Issues

Information Development through the 5 Phases of MIKE2.0

Continuous Implementation Phases

Activity 3.11 Data Profiling

Task 3.11.1 Prepare for Assessment
Task 3.11.2 Perform Column Profiling
Task 3.11.3 Perform Table Profiling
Task 3.11.4 Perform Multi-Table Profiling
Task 3.11.5 Finalize Data Quality Report

Phase 3 – Information Management Roadmap and Foundation Activities

<table>
<thead>
<tr>
<th>Activity</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1 Information Management Roadmap Overview</td>
<td></td>
</tr>
<tr>
<td>3.2 Testing and Deployment Plan</td>
<td></td>
</tr>
<tr>
<td>3.3 Software Development Readiness</td>
<td></td>
</tr>
<tr>
<td>3.4 Detailed Release Requirements</td>
<td></td>
</tr>
<tr>
<td>3.5 Business Scope for Improved Data Governance</td>
<td></td>
</tr>
<tr>
<td>3.6 Enterprise Information Architecture</td>
<td></td>
</tr>
<tr>
<td>3.7 Root Cause Analysis on Data Governance Issues</td>
<td></td>
</tr>
<tr>
<td>3.8 Data Governance Metrics</td>
<td></td>
</tr>
<tr>
<td>3.9 Database Design</td>
<td></td>
</tr>
<tr>
<td>3.10 Message Modeling</td>
<td></td>
</tr>
<tr>
<td>3.11 Data Profiling</td>
<td></td>
</tr>
<tr>
<td>3.12 Data Re-Engineering</td>
<td></td>
</tr>
<tr>
<td>3.13 Business Intelligence Initial Design and Prototype</td>
<td></td>
</tr>
<tr>
<td>3.14 Solution Architecture Definition/Revision</td>
<td></td>
</tr>
</tbody>
</table>

© 2008 BearingPoint, Inc.
The MIKE2.0 Methodology
Data Re-Engineering: Systematically Address Issues

Information Development through the 5 Phases of MIKE2.0

Continuous Implementation Phases

Strategic Programme Blueprint is done once

Phase 1 Business Assessment
Phase 2 Technology Assessment

Phase 3 – Information Management Roadmap and Foundation Activities

<table>
<thead>
<tr>
<th>Activity 3.12 Data Re-Engineering</th>
</tr>
</thead>
<tbody>
<tr>
<td>Task 3.12.1 Prepare for Re-Engineering</td>
</tr>
<tr>
<td>Task 3.12.2 Perform Data Standardisation</td>
</tr>
<tr>
<td>Task 3.12.3 Perform Data Correction</td>
</tr>
<tr>
<td>Task 3.12.4 Perform Data Matching and Consolidation</td>
</tr>
<tr>
<td>Task 3.12.5 Perform Data Enrichment</td>
</tr>
<tr>
<td>Task 3.12.6 Finalize Business Summary of Data Quality Impacts</td>
</tr>
</tbody>
</table>

© 2008 BearingPoint, Inc.
Implementing a Data Governance Programme
Guiding Principles

Achieve alignment of information models with business models from the get-go
- Enable people with the right skills to build and manage new information systems
- Improve processes around information compliance, policies, practices and measurement
- Create a culture of information excellence and information management organization structured to yield results

Deliver solutions that meet the needs of today's highly federated organizations
- Quantitatively identify data governance problems and resolve them
- Perform root cause analysis that led to poor data governance
- Remove complexity by ensuring all information is exchanged through standards
- Increase automation for the exchange of information across systems

Advance to a model focused on information development, just as we have models for developing applications and infrastructures
The Impact of Poor Data Quality
What are Some Lessons Learned?

**Drive Incremental Value Proportional to Business Demand**
- Align the Data Governance programme with a key business initiative that will contribute to your strategic goals
- Don't try to fix everything at once. Score quick wins along the way

**Take a Standards Based Approach to Data Governance**
- Data standards are the cornerstone of an effective Data Governance programme
- Balance centrally agreed standards vs. the need for a dynamic business with departmentally-driven goals
- Applications come and go, but the data largely stays the same. The Data Governance decisions you make today will have a profound impact on your business

**Take a Rigorous Approach to Organizational Alignment**
- Data Governance is not an IT-only initiative. It requires active involvement and leadership from the business as well as within the IT
- Am Executive Sponsor must provide leadership and senior data stewards must accept accountability
- Building an integrated and accurate enterprise view won't happen on its own – align initiatives across business units
Customer Data Integration and Data Quality – Architecture

**Background**

The Commonwealth Bank is one of Australia's leading providers of integrated financial services including retail, business and institutional banking, funds management, superannuation, insurance, investment and broking services.

The Commonwealth Bank is one of largest banks in the Australian financial services industry. In terms of scale, the Group has the largest customer base of any Australian bank and the most comprehensive financial services distribution network in Australia.

**The Business Challenge**

The Commonwealth Bank (CBA) had numerous customer information repositories representing holders of different products and customers of various business units. Obtaining a consolidated view of customers and their product holdings for analytical or operational purposes was extremely difficult. Do the lack of a 'single view of customer', the Bank believed that significant cross-selling and customer service opportunities were being lost. CBA wanted to implement information infrastructure that would facilitate the consolidation of customer information into the Bank's major analytical and operational data stores, thus allowing the leveraging of this information for business advantage.

**The Solution**

BearingPoint was engaged in partnership EDS, the Bank's outsourced IT provider, to implement major components of the customer information infrastructure, including:

- The Customer Matching Engine (Ascential's QualityStage tool)
- Processes to load customer data from six product systems. Development of this capability required the deployment of advanced data management principles, and adherence to the Bank's stringent compliance requirements. In addition it was required to comply with standard technology platforms. The scope of work included:
  - Building on-going capability to standardize customer data received from product systems;
  - Validation of data and appropriate reporting back to product systems;
  - Matching customers across product systems and the major operational database (CIF) to achieve single view of customer; and
  - Update and synchronization of operational databases and the analytical data warehouse.

**The Benefits**

BearingPoint’s contribution was key to the successful project implementation and project outcomes. This large project was implemented on time and within budget and resulted in delivering a consolidated 'single view of customer' and customer product holdings for six product systems, facilitating numerous business benefits. This project is seen as pioneering in this sphere and the bank is now in the process of planning to introduce data from several other product systems into this core customer infrastructure.

© 2008 BearingPoint, Inc.
Customer Data Integration and Data Quality – Architecture

Customer Staging Area

- Customer Oriented Data
  - CIF
  - RDB
  - Operational Deltas
  - Initial Loads
  - Test Loads
  - Standardize
  - Validate
  - De-dupe
  - Match
  - Common Interface
  - DWE Updates
  - Exception Reports
  - Data for manual updates

Account Oriented Data

- Data Loads
  - CBI
  - CBL
  - BCC
  - CIL
  - CBFCCL
  - CBFCFD

Customer-based Operational Systems

© 2008 BearingPoint, Inc.
Customer Data Integration and Whole of Customer View

**Background**
The Commonwealth Bank is one of Australia's leading providers of integrated financial services including retail, business and institutional banking, funds management, superannuation, insurance, investment and broking services.
The Commonwealth Bank is one of the largest banks in the Australian financial services industry. In terms of scale, the Group has the largest customer base of any Australian bank and the most comprehensive financial services distribution network in Australia.

**The Business Challenge**
The objective of the project was to develop Data Loads of customer data from key product source systems for loading to the Bank's key operational and analytical data stores; the Customer Information Facility (CIF) and the Group Data Warehouse (GDW). This data, when combined with data loaded for six other source systems previously, provides a consolidated view of customers and their product holdings to enable cross-system analysis and reporting. This project leveraged existing infrastructure and data load processes from a previous project, DLMA Phase One, which shared the same objective of enabling the SCV. The Customer Matching Engine and common Data Load processes built during DLMA phase one was leveraged and reused where appropriate.

**The Solution**
BearingPoint was engaged to work in partnership with EDS on the strength of our key Data Management expertise and experience in developing the core infrastructure during the previous Data Loads Matching Application project. The key activities in the project undertaken by the joint CBA, EDS and BearingPoint team included:

- Data Profiling – Investigation and assessment of data quality for the selected source systems using Ascential's Integrity, ProfileStage and AuditStage
- Data Mapping – Development of the business rules and requirements for processing the data using Ascential's DataStage
- Recommendations for on-going data quality improvement
- Development, Testing and Deployment of Data Loads – Creation of the data load programs using a combination of ETI, COBOL, Easytrieve and Ascential's Integrity

**The Benefits**
The Core Customer Information (CCI) initiative originated with a primary objective of providing central infrastructure for delivering core customer information to support a 'single view of customer' (SCV) throughout the Bank. Following the implementation of CCI Customer Data Loads Release One, the coverage of customers in the Bank's core customer repositories has been improved further enabling the various initiatives in the Bank that leverage this data. CCI Customer Data Loads Release Two is currently in progress, with an implementation date of April 2005.
Data Quality Stream for Single Customer View

Background
Insurance Australia Group is the largest general insurance group in Australia and New Zealand. It provides personal and commercial insurance products under some of the most respected and trusted brands including NRMA Insurance, SGIO, SGIC, CGU, Swann Insurance, and State. Insurance Australia Group's core lines of insurance business include:
- Home
- Health
- Extended warranty
- Compulsory third party
- Motor vehicle
- Commercial
- Consumer Credit
- Workers' compensation

The Business Challenge
IAG's challenge was to address serious deficiencies in the data quality of their main insurance system, particularly around client identification, de-duplication and achieving a view of the customer's entire list of business (lob). BearingPoint were engaged to develop a new match engine to improve de-duplication of the customer database, provide QA on data analysis and cleansing activities and to conduct a Business Impact Analysis (BIA) on the implementation of new business rules. As a result the match rate was improved by approx. 7%, rules around matching and cross-population of data (survivorship) developed and impact assessed and a clerical review process established for potential over/under match results.

The Solution
Key aspects of BearingPoint's solution included:
- Project Quality Assurance, providing IAG with guidance around DQ analysis, cleansing and establishment of DQ metrics & standards
- Development of a new customer match engine to de-duplicate the customer database
- Business Impact Analysis to develop business rules around matching and survivorship and assess the potential impacts of implementing those rules from financial, non-financial and privacy perspectives
- A Data Quality Assessment of three core insurance systems to identify major issues and assist in developing of a data cleansing strategy
- Development of a post-match process to enable a clerical review of potential over/under matches

The Benefits
The project has improved the process for identifying duplicate records, reduced the risk of under/over matching and ensured project methodologies and solutions include data quality issues and solutions in the design. The end result of the engagement will ensure data quality is improved, match rate increased, risks identified, business impacts assessed and controlled and matching processes improved. Client de-duplication rate was improved by approximately 7% over existing process and external auditor endorsed the process and results of the new matching and post-match processes. A Business rules repository established for matching and survivorship and materiality of any potential business impacts by implementation of new rules quantified. As a whole, this established a baseline of greatly improved data quality and customer lob for the overall SCV solution.

© 2008 BearingPoint, Inc.
Supply Chain Improvement Data Quality Strategy and Implementation

Background
The Defence Materiel Organization (DMO) is responsible for acquisition of capital equipment and systems and through-life support of materiel for the Australian Defence Force. DMO is a large, complex and distributed organization that supports assets worth $38b, has over 8000 staff and operates from some 76 locations in Australia and overseas.

A business unit within the DMO’s Materiel Systems Branch, Supply Chain System Program Office (SPO) is responsible for the through life support of the Standard Defence Supply System (SDSS) applications and interfaces.

The Business Challenge
DMO has in recent years faced increasing scrutiny and attention with regards to the management of Assets and regulatory reporting of financial information to the Government.

The lack of internal controls in the prime logistics system, SDSS, has been a major cause for concern for the Department and has resulted in ongoing qualifications of its financial accounts by the Australian National Audit Office (ANAO).

The Solution

- BearingPoint were engaged to assist the DMO – Materiel Systems Branch with the formulation of a Data Quality (DQ) Strategies and the implementation of a Data Management Framework to provide a holistic approach towards Data Management in SDSS.
- BearingPoint worked closely with an internal Defence DQ team to develop and implement a number of Data Quality strategies using a methodology to address known Data issues based upon leading Data Management processes.
- Specifically, BearingPoint designed and implemented a DQ Initiative to address the data deficiency of Supply Customer accounts in SDSS. This included analysis, investigation and remediation of missing and invalid data as well as the development of a monitoring program to perform ongoing ‘health checks’ of Data quality.
- BearingPoint led the design, development and implementation of the DQ Pricing remediation initiative in order to establish the baseline of valid Logistic Stock items which have a quantifiable price and source in SDSS.

The Benefits
As a result, BearingPoint was able to provide DMO with a number of Data quality deliverables and work packages based upon leading Data Management practices. BearingPoint developed and implemented a methodology for Data Quality monitoring and reporting which provided the client with assurance for maintaining Data Quality levels in the organization for the specific business areas of Supply Customers and Stock Item Pricing.
Customer Data Integration and Data Quality Improvements

Background
Telstra Corporation is Australia’s largest telecommunications provider. 36,000 staff members work to offer a full range of services throughout Australia, providing more than 10.3 million Australian fixed line and more than 6.5 million mobile services. The company announced a $4.1 billion profit for the 2004 financial year.

The Business Challenge
The project was established to determine monthly "net new" customers from the pay-TV reseller initiative between Foxtel and Telstra (Project Mulder). The major objective of the "Net New" Matching Project was to provide an accurate and auditable mechanism to determine new customers versus existing Foxtel customers. This would allow Telstra to calculate commission payments each month.

The Solution
The "Net New" solution was built around Ascential Software's QualityStage data matching tool. The features of the solution include:
- Fuzzy-logic matching of records
- Automatic end-to-end development of match results from raw input files controlled by scripts
- Email notifications and event management
- Metadata management within and outside Integrity
- Industry-proven database for storing match results and performing calculations
- Log files serve as an audit trail of events
- Standard Integrity reports
- A full backup and storage facility of monthly results

The Benefits
The "Net New" Matching Project was completed successfully in January 2003, giving Telstra the capability to confidently calculate commission through the Foxtel reseller agreement. The system is easy-to-use, robust and highly automated for finding "net new" customers with quick results.
Customer Data Integration and Data Quality Improvements

Wizard

Major functions include:
- Disputation
- Matching to Baseline
- Auditing

Redhead

Technologies include: Integrity, Oracle, Perl, Solaris, FTP, TCP/IP
Data Quality Strategy and Implementation

**Background**
Union Bank of California is the fourth-largest commercial bank in California and one of the 25 largest banks in the United States. The bank has 316 banking offices throughout California, Oregon, and Washington, as well as offices in New York and Texas. It also has 21 international offices, mainly along the Pacific Rim.

**The Business Challenge**
The Bank has long recognized the need to manage the quality of its data. As part of the effort in building an Enterprise Data Warehouse, the Bank desired to implement an ongoing data quality process at the enterprise level. UBOC asked BearingPoint provide an assessment of their data quality issues and a recommended improvement plan.

**Solution**
- **Data Quality Assessment:**
  - A Tool-based assessment was conducted using Ascential's product suite. This involved source system data profiling, business rules testing, and customer record duplication identification.
  - Unlike most data quality assessments, meaningful business rules were tested for consistency across systems.
  - A comprehensive data quality survey was administered to executives representing the full enterprise.
- **Recommended Improvement Plan:**
  - Three-phased improvement roadmap, incrementally achieving higher information management maturity levels.
  - Conceptual architecture that represents long-term direction. Many short-term steps will be achieved along the way.
  - Business case that justifies the investment.

**The Benefits**
BearingPoint conducted a survey that identified the gap between current and target information maturity levels as the roadmap basis. The tool-based assessment results were used to adjust priorities, identify Quick Win opportunities, and to serve as a basis for a metadata repository. BearingPoint's Information Maturity Framework was used as to structure the data quality improvement analysis.
Data Integration & DPID Matching

**Background**
Sensis is a leading directory, advertising and information provider. Sensis offers a multi-channel solutions portfolio that incorporates advertising and information across directories, voice, online and wireless channels and electronic directories. Sensis manages some of Australia's most iconic brands, such as: White Pages, Yellow Pages, CitySearch and "Whereis".

**The Business Challenge**
The objective of the project was to deliver an in-house data integration and matching solution that would provide Sensis with the capability to standardize customer address data and match Delivery Point Identifiers (DPID's) to addresses for mailing collateral. DPID's are bar codes that identify each address in Australia and are used by Australia Post for the efficient delivery of mail. Sensis required the use of DPID matching to improve its success rate in delivering correspondence to customers and to obtain the substantial discount offered by Australia Post for the use of DPID's.

**The Solution**
- BearingPoint deployed a solution built around Ascential Software's DataStage extract, transform and load (ETL) tool and QualityStage data standardisation and matching application. This solution is fully automated and was one of the first implementations in Australia of the QualityStage DPID matching module.
- DataStage is used to pull files of customer correspondence, billing and direct marketing mail-out lists from a Sensis staging area, transform them and pass them to QualityStage for DPID matching. QualityStage standardizes the data in the files and appends the DPID to the customer addresses. DataStage forwards the updated data files to an external mail house for addressing and sending correspondence to customers. Full error handling and email notification of ETL job success were deployed as part of the solution.
- BearingPoint developed a reporting capability that enables Sensis to determine which customer addresses are failing the DPID append process so that corrections can be made to improve customer contact address data.

**The Benefits**
The data integration and DPID matching project was delivered successfully in 1Q04 and is currently being used to drive down the costs associated with client contact initiatives. Subsequent phases of work will leverage the solution to enable a "Whole of Customer View" that aggregates customer information across multiple legacy systems prior to loading data into the Sensis' enterprise data warehouse for reporting purposes.
Data Integration & DPID Matching

Source Systems

- Prism/BCV (Genesis Mirror)
- Sensis Data Warehouse
  - FTP/SMTPl Manual Download/Email
  - SCP Secure Copy of Billing Files

FTP Servers/Share Folders

- [PAUX02 Server]
  - Daily CC Packs – Orig
  - Daily CC Packs – Clean
  - Daily Billing File – Orig
  - Daily Billing File – Clean
  - Ad-hoc DM Mail Lists – Orig
  - Ad-hoc DM Mail Lists – Clean

FTP/SCP of CC Packs

Data Quality Environment

- Telstra Co-location Data Centre
- Data Cleansing & Integration
  - Daily Billing File DQ Process
  - Daily CC Packs – DQ Process
  - Ad-hoc DM Lists DQ Process

- DQDB502P Server
  - Data Stage
  - Data Cleansing
  - Quality Stage

Legend

- Original Files in Shared Folder
- Cleansed Files in Shared Folder
- Processing Files in DQ Environment
- Data Quality Software Tools

Sensis Work

- Manual FTP Download of DM Mail Lists
- Compile DataStage & QualityStage Jobs
- Create DataStage & QualityStage Jobs
- Data Quality Client Tester workstation
- Data Quality Client Developer Workstation
- Data Quality Analyst
- IT Support Help-Desk

Exception Reports delivered by SMTP
Data Quality Job Status delivered by SMTP

© 2008 BearingPoint, Inc.