

7th International Conference on Information Quality (IQ-2002)

DATA QUALITY IN ACTION: CHALLENGE IN AN INSURANCE COMPANY

Miguel Feldens
GoDigital Precision Marketing
miguel@godigital.com.br

Dalvani Lima
GoDigital Precision Marketing
dalvani@godigital.com.br

Luis Francisco Lima
GoDigital Precision Marketing
francisco@godigital.com.br

Executive Summary/Abstract: It is obvious to say that data quality management must be a major concern in all industries. Strategies for customer relationship management (CRM), and business intelligence (BI), for example, can only be as good as the quality of data. Both internal and external data stores should have quality standards defined, measured, analyzed, and improved, to succeed.

Insurance companies are challenging to the implementation of data quality management. This presentation is the result of a project at Cosep, a large insurance company in Brazil, and covers, as much as can be disclosed, several issues of real world data quality management implementations.

7th International Conference on Information Quality (IQ-2002)

Objectives of this presentation

- **Contribute** to the research on Information Quality (IQ), providing information about a real world implementation in a challenging scenario
- **Analyze** this scenario under the perspective of Total Data Quality Management (TDQM)
- **Identify** contributions of TDQM to the insurance business
- **Move towards** a replicable methodology for TDQM in insurance

- Poor data quality
- Demanding Information Product (IP) consumers
- Implications of value chain, culture and technology on IQ
- Identify IP consumers, suppliers, manufacturers and managers
- Analytical and operational applications of qualified data

7th International Conference on Information Quality (IQ-2002)

Contents

Scenario

1. Value chain, culture, and technology
2. IP suppliers and Quality Evaluation
3. Corrupted Information: an Example

Methodology: Unique Customer View (UCV)

1. Measuring and Analyzing Data Quality
2. Data Reengineering
3. Key Performance Indicators (KPIs)

Results

1. Fraud / Anomalies Detection
2. Analytical Results & Opportunities

Further work

1. What's next?

7th International Conference on Information Quality (IQ-2002)

Scenario

Brazilian Insurance Value Chain:
Insurance companies cannot sell directly to the customer. All sales must be intermediated by brokers

Due to regulatory laws the "Sales Channel" must be a Retail Bank, specialized Insurance Broker, or other financial Services Company

Channels

- Provide "insurance solution" to the customer through "one window"
- Allow the customer to choose among products from different insurance companies
- Increase competitiveness between insurance companies
- "Own" (sales) relationship with the customer (and related data)

Sales Relationships

7th International Conference on Information Quality (IQ-2002)

Scenario

Brazilian Insurance Value Chain:
Claims, assistances, refunds covered by insurance are generally attended through specialized assistance companies

Most insurance companies choose to delegate customer care, assistance, and other insured services / coverage to third-parties

Assistance companies

- Manage phone calls
- Support emergency services, such as tow-cars, repairs, ambulances
- Support filling claims and directs them to the insurance company
- Keep databases on assistances, and replicate to the insurance company when necessary

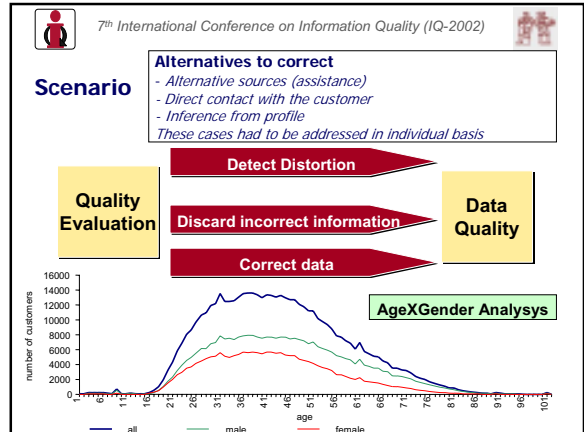
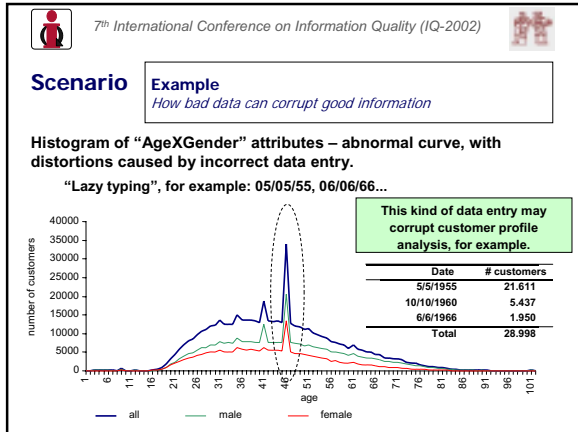
Customer Assistance Relationships

7th International Conference on Information Quality (IQ-2002)

Scenario

Data Quality Assessment:
Highlights from data audit and interviews at Cosep - Data quality problems due to "intermediated" relationships

IQ Categories and Dimensions [3]		Insurance Scenario
IQ Category	IQ Dimensions	
Intrinsic IQ	Accuracy, Objectivity, Believability, Reputation	Only essential information for pricing is reliable, but differs on each product (auto and life insurances have different attributes). There are multiple representations of the same customer in the database, as he/she can buy from different banks.
Accessibility IQ	Access, Security	All data that is needed for operational processes (including data from channels and assistance) are replicated into insurance company mainframe.
Contextual IQ	Relevancy, Value-Added, Timeliness, Completeness, Amount of data	Customer demographics from the deal closing may be outdated. More recent data can be collected from the assistances.
Representational IQ	Interpretability, Ease of understanding, Concise representation, Consistent representation	Different assistance companies may support the insurance company and use their own naming and codes for assistances and services.



7th International Conference on Information Quality (IQ-2002)

Contents

Scenario

1. Value chain, culture, and technology
2. IP suppliers and Quality Evaluation
3. Corrupted Information: an Example

Methodology: Unique Customer View (UCV)

1. Measuring and Analyzing Data Quality
2. Data Reengineering
3. Key Performance Indicators (KPIs)

Results

1. Fraud / Anomalies Detection
2. Analytical Results & Opportunities

Further work

1. What's next?

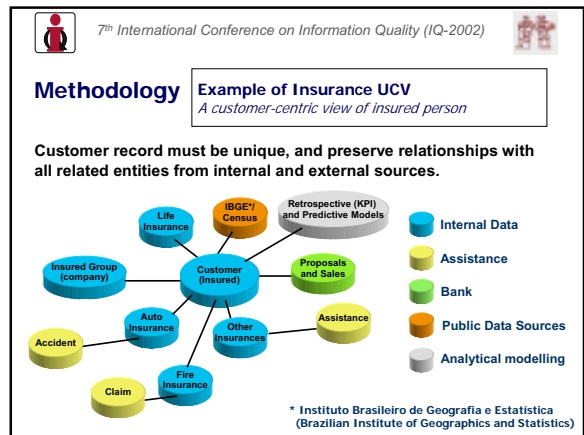
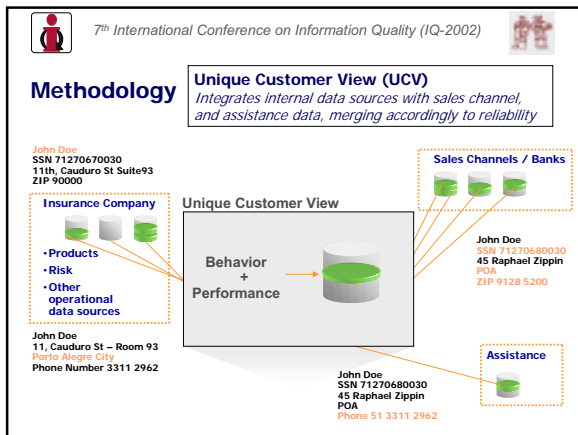
7th International Conference on Information Quality (IQ-2002)

Methodology Overview & Results

Data Stores Vs. Unique Customer View

Data has been qualified, modeled, and enriched to support relationship management and decision making.

	Cosep's Operational Data Sources (original)	Unique Customer View (result of this project)
Goal	• Operation	• 1-to-1 relationship & decision
Data Model	• Lacking / inconsistent integrity constraints • Poor documentation	• Standard representations • Validation • Complete documentation
Quality	• Duplicate records > 20% • Without gender = 90% • 76% of addresses incomplete or incorrect	• Duplicate records < 5% • Without gender < 13% • 67% of addresses completed or corrected
Unique Customer Identification	• Not Available	• Cleaned, and enriched with assistance and sales data • Linking to all claims, proofs of insurance, assistances...



7th International Conference on Information Quality (IQ-2002)

Methodology

Steps Towards UCV in Insurance

Once IQ of internal data stores is measured and analysed, there are four steps to build the UCV

1 + 2 + 3 + 4

Structure Customer View

Understanding IP consumer needs, and business strategies.

Modelling the necessary data structures to support them.

Information Integration

Mapping different schemas, building and propagating keys to link external data sources.

Extract, transform, and load operational data, into the UCV.

Quality Improvement

Merge-and-purge. Enrichment with external data. Inference, and validation.

Modelling key performance indicators (KPIs) and strategic information for each customer, channel, and product

- Retrospective
 - Total revenue
 - Total accidents
 - Total cost of sales
 - Total cost to serve
 - Contribution margin and profit
- Predictive (data mining and scoring)
 - Buying power
 - Next best offer

Unique Customer View

7th International Conference on Information Quality (IQ-2002)

Methodology

Structure Customer Unique View

Conceptual model and quality requirements to support strategy and decision making

1

Transactions (proofs of insurance)
Channel Value
Mean of Payment

Insurance (products)

Payment ...

Accident ...

Coverage ...

Refund ...

Assisnancy ...

Beneficiary ...

Behavior
Channel
Event
Campaign

Insured
Unique ID

Demographics ...

Behavior & KPIs
Recency
Frequency
Value
Channels
Products
Credit risk
Churn risk
Accident risk

Conceptual model, including behavior and performance indexes.

Quality requirements were also modeled for all these information [2]

7th International Conference on Information Quality (IQ-2002)

Methodology

Information Integration

Implementation of ETL (extract, transform, and load) to populate UCV and related entities

2

Insurance Company

Sales Channels / Banks

Unique Customer View

→ **Transactions**

→ **Insured**

→ **Assisnances**

... plus other information to complete the model, generating a data set able to fit the needs for further use.

7th International Conference on Information Quality (IQ-2002)

Methodology

Quality Improvement

Supporting IQ with Go Digital's Go Quality

3

Software support for quality requirement definition, corrective data quality (cleansing) and real-time data quality management.

→ **Standardization / Cleansing**

- Abbreviations, prefixes and suffixes
- Addresses (ZIPs, cities, states, streets)
- Phone numbers (intern. and local codes)
- Gender inference
- Company / person name detection
- Dirty word detection
- House / businesshold differentiation
- URL / Email syntax validation

→ **Enrichment** (Statistics, Post office databases, Research data)

→ **Deduplication** supported with multiple weighted matchcodes, approximate matching, and phonetic algorithms

→ **Data Quality operations and standards** available through API (application programming interface)

7th International Conference on Information Quality (IQ-2002)

Methodology

Predictive Models

Predictive models built with data mining, and applied to the whole customer database

4

Unique Customer View

Customers

Go Mining

Insured Unique ID

Demographics ...

Behavior & KPIs

- Recency
- Frequency
- Value
- Channels
- Products
- Next best offer
- Credit risk
- Churn risk
- Accident risk

7th International Conference on Information Quality (IQ-2002)

Methodology

KPIs and Predictive Models

Consistent, unique, and enriched data support analytical processing

Performance per customer, per product and per channel

Accident Costs (%)

Category	Goods	Home	Auto	Civil Resp
Value	0	9.3	5.5	23.2

Revenue (%)

Category	Goods	Home	Auto	Civil Resp
Value	100	100	100	100

Sales Costs (%)

Category	Goods	Home	Auto	Civil Resp
Value	6.2	0.0	7.8	9.6

General and administrative Costs (%)

Category	Goods	Home	Auto	Civil Resp
Value	8.2	9.3	13.2	32.8

Combined Index (%)

Category	Goods	Home	Auto	Civil Resp
Value	93.85	90.7	95.8	67.2

Result (%)

Category	Goods	Home	Auto	Civil Resp
Value	93.85	90.7	95.8	67.2

* Products and numbers changed for privacy
Costs not shown according to non-disclosure agreement

7th International Conference on Information Quality (IQ-2002)

Contents

Scenario

1. Value chain, culture, and technology
2. IP suppliers and Quality Evaluation
3. Corrupted Information: an Example

Methodology: Unique Customer View (UCV)

1. Measuring and Analyzing Data Quality
2. Data Reengineering
3. Key Performance Indicators (KPIs)

Results

1. Fraud / Anomalies Detection
2. Analytical Results & Opportunities

Further work

1. What's next?

7th International Conference on Information Quality (IQ-2002)

Results

Unique Customer View = Information Asset
Integrating disperse information and creating new attributes from existing data, creating a valuable information asset.

Customer
 - SSN 71270670030
 - John Doe

Transactions
 - SSN 71270670030
 - payment \$9090.00
 - Health Insurance

Assistance
 - John Doe
 - SSN 71270670030
 - Hospital
 - Heart disease

Census Information
 - Sao Paulo city
 - ZIP 9128 5200
 - Heart disease tax: 30%
 - % male: 65%

Unique View
 - SSN 71270670030
 - John Doe
 - Sao Paulo city
 - ZIP 9128 5200
 - gender: male
 - risk: 30%
 - # assistances: 1
 - payment average: \$90

Highlights

Hundreds of small programs running on, requesting and modifying data from more than 300 tables in a mainframe architecture, with disperse information

A UCV has been consolidated in this company, using the Data Quality methodology which is being presented.

7th International Conference on Information Quality (IQ-2002)

Results

Increased Analytical Power
 Abnormal behaviors and outliers in the KPIs raise "red flags", and can reveal potential frauds

Some actual discoveries:

- Regions – a specific state had very poor results
 - Distortion in policies adopted for that particular state
- Insured good – specific brand of auto with very high rate of accidents
 - Potential fraud or anomaly in a large fleet of cabs
- ...

7th International Conference on Information Quality (IQ-2002)

Results

Increased Analytical Power
Treatment and enrichment of customer data plus Data Mining, revealing more than US \$ 10 M in cross-sales opportunities just for the active customer base.

- Discovery of more than 100 business rules with data mining
 - Customer demographics & current products → future purchase
- Assessment of active customers using discovered rules:

Product*	# Opportunities	Avg Price (\$)	Addressable (\$)
Insurance 1	12.823	383,81	4.921.595,63
Insurance 2	3.618	1.253,52	4.535.235,36
Insurance 3	8.073	232,28	1.875.196,44
Insurance 4	2.155	-	-
Insurance 5	1.048	-	-

* Actual products and numbers where changed due to non-disclosure agreement

7th International Conference on Information Quality (IQ-2002)

Contents

Scenario

1. Value chain, culture, and technology
2. IP suppliers and Quality Evaluation
3. Corrupted Information: an Example

Methodology: Unique Customer View (UCV)

1. Measuring and Analyzing Data Quality
2. Data Reengineering
3. Key Performance Indicators (KPIs)

Results

1. Fraud / Anomalies Detection
2. Analytical Results & Opportunities

Further work

1. What's next?

7th International Conference on Information Quality (IQ-2002)

Further Work

Next Steps
Assessment, measurement, analysis, cleansing, and enrichment were large steps, but it is just the beginning... More value could be added through data quality management.

- On-line Data Quality
 - Integrate Data Quality operations with Assistance Call Center
 - Company website
 - Operational applications
- Value-added information in the touch-points with the customer
 - Integrate KPIs, predictive models, and differentiated policies
 - Call center
 - Website
 - APIs (application programming interfaces) available as services to banks and financial services partners



7th International Conference on Information Quality (IQ-2002)



References

- [1] Andersen. *Brazilian Insurance Market 2000/2001*. Andersen, 2001.
- [2] Eyer, T. et al. *Reinventing Insurance Sales*. The Forrester Report. Forrester Research, 2001.
- [3] Wang, R. Y. A "Product Perspective on Data Quality Management". *Communications of the ACM*, Vol. 41, No. 2, February 1998, p. 58-65