



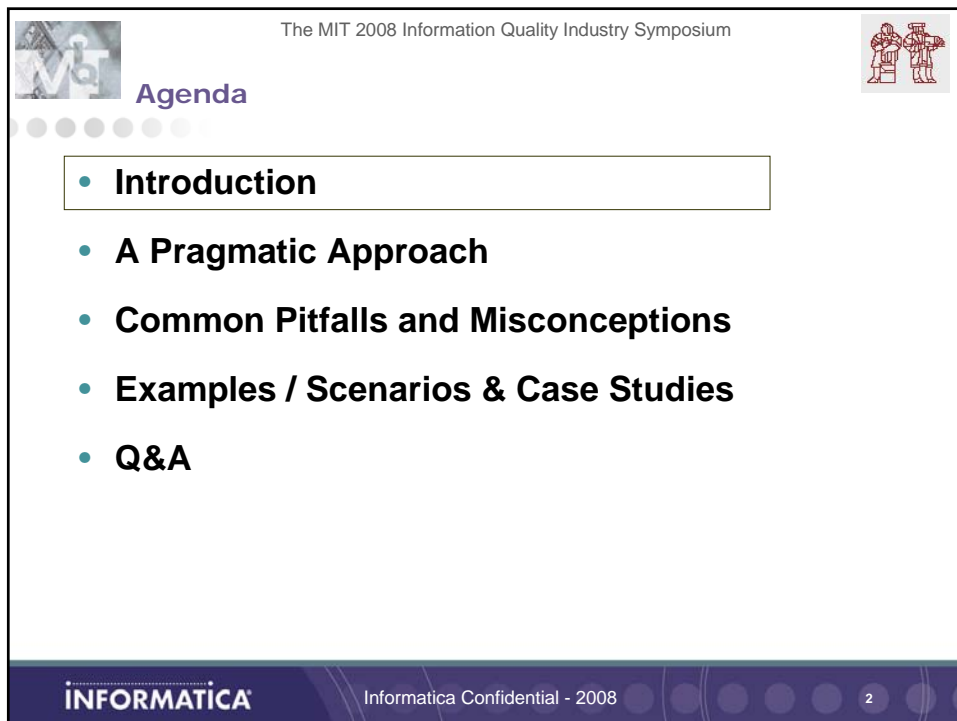
**INFORMATICA**

## *Data Quality for Successful Data Governance*


*Ivan Chong  
EVP and General Manager  
Data Quality Business Unit  
Informatica Corporation*

Informatica Confidential - 2008


1



The MIT 2008 Information Quality Industry Symposium



**Agenda**



- **Introduction**
- **A Pragmatic Approach**
- **Common Pitfalls and Misconceptions**
- **Examples / Scenarios & Case Studies**
- **Q&A**

**INFORMATICA**

Informatica Confidential - 2008

2

The MIT 2008 Information Quality Industry Symposium

## Introduction: Data Quality Evolution

The diagram illustrates the evolution of data quality solutions through four stages:

- Contact Efficiency:** Parsing, standardization, cleansing and validation of customer contact-related data elements, such as names, addresses and telephone numbers.
- Relationship Identification:** Discerning relationships between records (for example, householding, de-duplicating or linking).
- Customized Data Quality:** Data quality operations for a range of data subject areas, including product data and financial data. This stage includes **SCM** (Supply Chain Management).
- Data Quality Analysis:** Data profiling, measurement and quantification of data quality impact, and ongoing auditing and monitoring of data. This stage includes **PIM / MDM** (Product Information Management / Master Data Management).

Additional solutions associated with the Enterprise DQ Suite include **SOX**, **Six Sigma**, and **Compliance**.

**INFORMATICA** Informatica Confidential - 2008 3

The MIT 2008 Information Quality Industry Symposium

## Introduction: Impact of Poor Data Processes & Data Quality


Poor Data Quality or inconsistent data causes defects in the value chain and is a momentum killer for any Data Alignment, Synchronization and Collaboration project, Master Data Management, CDI, Application Implementation etc ...

The iceberg diagram illustrates the impact of poor data quality, with visible issues above the waterline and hidden costs below:

- Visible Issues (Above Waterline):**
  - Out of Stock Situations
  - Incorrect Orders
  - Delivery Errors
  - Pricing Discrepancies
  - Speed to Market
- Hidden Costs (Below Waterline):**
  - Difficulty in decision making
  - Penalties from large retailers
  - Invoice reconciliation issues
  - Time to introduce new trading partner
  - Surplus inventory
  - Customer service time
  - Logistics costs
  - Organizational mistrust
  - Inaccurate plan-o-grams
  - Lost customer loyalty

**INFORMATICA** Informatica Confidential - 2008 4

The MIT 2008 Information Quality Industry Symposium



## Introduction: One Perspective to Governance, top down

**Governing corporate data may be slightly easier than governing nations, but companies can benefit from adopting similar strategies, according to a Gartner analyst.**

**Executive level.** sponsorship, strategic direction, funding, advocacy and oversight

**Judicial level.** planning activities and to enforce governance activities or corporate policies. Mediating disagreements



**Legislative level.** chaired by a senior business leader designated by the executive team and may include business and technology leaders from Finance, IT, Data Management and Operations

**Administrative level.** Implement data governance on a day-to-day basis; responsible for developing data models and corporate data vocabularies, implementing master data management best practices, organizing content

**Top Down – often initiated by Exec initiative – Compliance / Audit / BPR / Six Sigma**

**INFORMATICA** Informatica Confidential - 2008 5

The MIT 2008 Information Quality Industry Symposium



## Introduction: Data Quality Perspective to Governance

Lack of a comprehensive **Data Quality Management Strategy** and poor **Data Quality** is causing **Master Data Management (MDM) initiatives to fail or be significantly delayed.**

**key elements of effective Data Quality Management:**

- capable data processes
- ownership and accountability
- standards and metrics
- measurement and control

**key stages in any migration or MDM implementation**



- discovery and assessment
- planning and preparation
- cleansing, alignment, enrichment and enhancement
- operational impact and support in preparation for, during and after implementation
- **Transition management**
- post implementation measurement and control

**Operational Risk**

**Bottom Up – often initiated by project – DQ / Migration / App Implement / Six Sigma**

**INFORMATICA** Informatica Confidential - 2008 6

The MIT 2008 Information Quality Industry Symposium





## Agenda

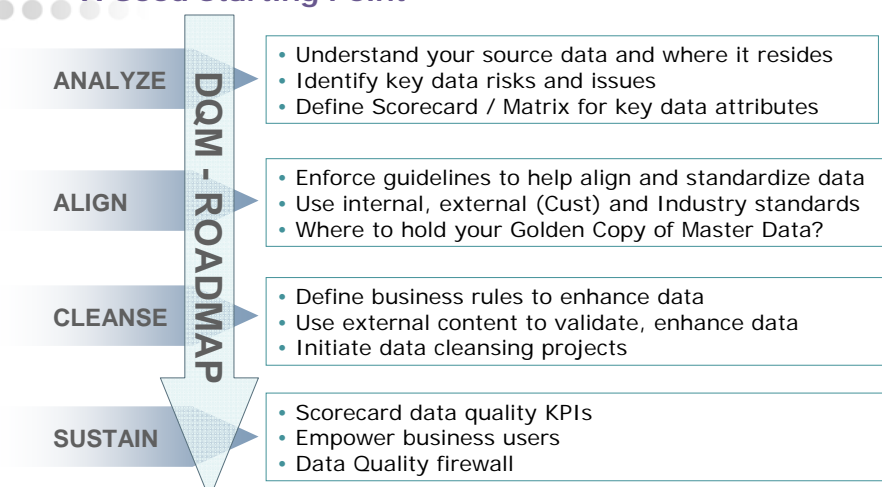
- Introduction
- **A Pragmatic Approach**
- Common Pitfalls and Misconceptions
- Examples / Scenarios & Case Studies
- Q&A

**INFORMATICA** Informatica Confidential - 2008 7

The MIT 2008 Information Quality Industry Symposium



## Data Quality Management – A Good Starting Point





**DQM - ROADMAP**

- ANALYZE**
  - Understand your source data and where it resides
  - Identify key data risks and issues
  - Define Scorecard / Matrix for key data attributes
- ALIGN**
  - Enforce guidelines to help align and standardize data
  - Use internal, external (Cust) and Industry standards
  - Where to hold your Golden Copy of Master Data?
- CLEANSE**
  - Define business rules to enhance data
  - Use external content to validate, enhance data
  - Initiate data cleansing projects
- SUSTAIN**
  - Scorecard data quality KPIs
  - Empower business users
  - Data Quality firewall

**INFORMATICA** Informatica Confidential - 2008 8

The MIT 2008 Information Quality Industry Symposium






## Agenda

- Introduction
- A Pragmatic Approach
- **Common Pitfalls and Misconceptions**
- Examples / Scenarios & Case Studies
- Q&A

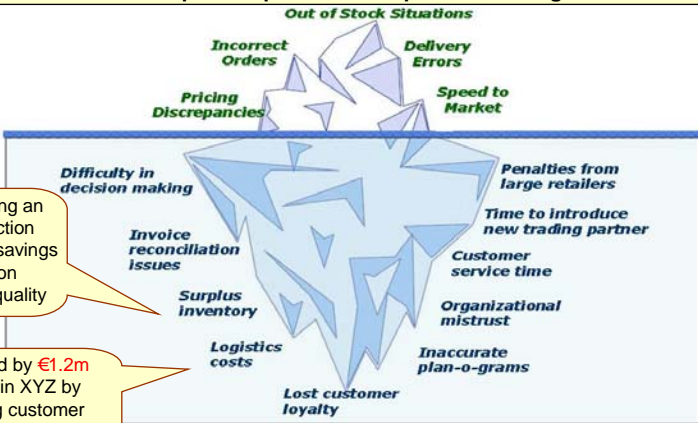
**INFORMATICA** Informatica Confidential - 2008 9

The MIT 2008 Information Quality Industry Symposium

## Background: Focus on value early

The impacts of poor quality data are not always highly visible. Using Supply Chain as an example here, many impacts are hidden 'below the waterline'. The visible impacts represent the 'tip of the iceberg'



When considering an inventory reduction initiative, 13% of savings were based on improving data quality

Decreased by €1.2m annually in XYZ by improving customer delivery time data

**INFORMATICA** Informatica Confidential - 2008 10

The MIT 2008 Information Quality Industry Symposium

## Information Governance Framework

**Be clear of what is in scope for 'DATA GOVERNANCE'**

**Information Governance**

- Data Protection
- Licensed 3<sup>rd</sup> PTY Data
- Legislation

**Data Governance**

- Business Processes
- Data Quality
- Roles and Response
- Business context

**Technology Governance**

- Data Security
- Anti-virus
- Disaster Recovery

**INFORMATICA** Informatica Confidential - 2008 11

The MIT 2008 Information Quality Industry Symposium

## MDM – Generic concerns

**Figure 1 The MDM Ecosystem Audit Guide**

**MDM Applications / Solutions**

- Decouple Master and Reference Data capture from Legacy
- All Master Data captured in one application process across the enterprise
- Operational Data Stores and Data Warehouse Risks


**Implementation Risks**

- Registry / Repository / Hub
- Focus on quick wins and value
- **Operational Risk, Internal & External**

41343 Source: Forrester Research, Inc.

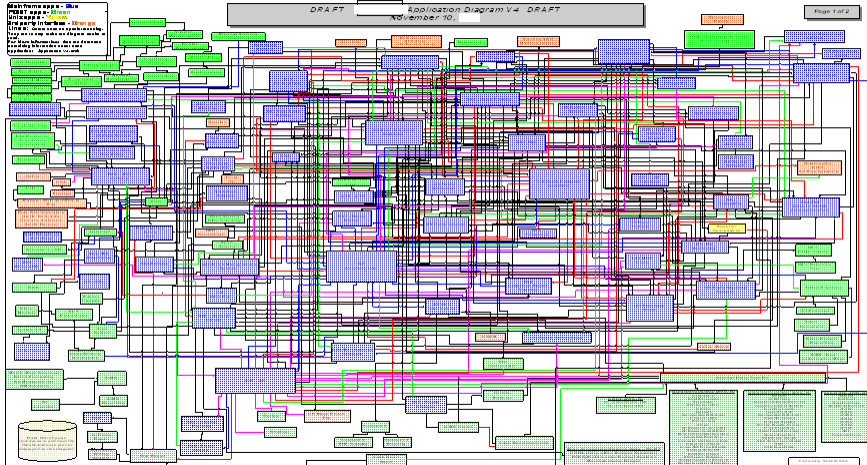
**INFORMATICA** Informatica Confidential - 2008 12

The MIT 2008 Information Quality Industry Symposium



# The Challenge

## Data Integration Complexity




INFORMATICA

Informatica Confidential - 2008

13


The MIT 2008 Information Quality Industry Symposium



# The Challenge

## Early Analysis is critical – Obsolescence / Duplicates

**Data Alignment means focus on Data Quality**



Results of worldwide Data Cleansing:

	Obsolete	Duplicate	Active	Cleansing rate
Material	724'804	29'529	544'383	58%
Customer	1'728'298	105'033	1'733'809	51%
Vendor	1'079'660	23'266	632'952	64%
<b>Total</b>	<b>3'532'762</b>	<b>157'828</b>	<b>2'911'144</b>	<b>56%</b>

Cleansing includes completeness and correctness checks on active data records



INFORMATICA

Informatica Confidential - 2008

14




The MIT 2008 Information Quality Industry Symposium



## One Customers Experience - Large US Bank

---





### Lessons Learned

- Lessons Learned #1
  - Too much, too fast. Don't try to solve all the problems at once
  - Give the stakeholders enough to start with, but do not overload them
  - Only give the stakeholders the defects that apply to them & separate defects by type:
    - o Data Entry defects → Front line stakeholders
    - o Data Movement defects → Extract/Load Technology teams
    - o Data Enrichment → Transformation Technology teams
- Lessons Learned #2
  - Create data quality business rules that align with Policies and Procedures
  - Don't create rules that "seem to make sense". Create rules that are "Actionable"

**INFORMATICA** Informatica Confidential - 2008 15

The MIT 2008 Information Quality Industry Symposium

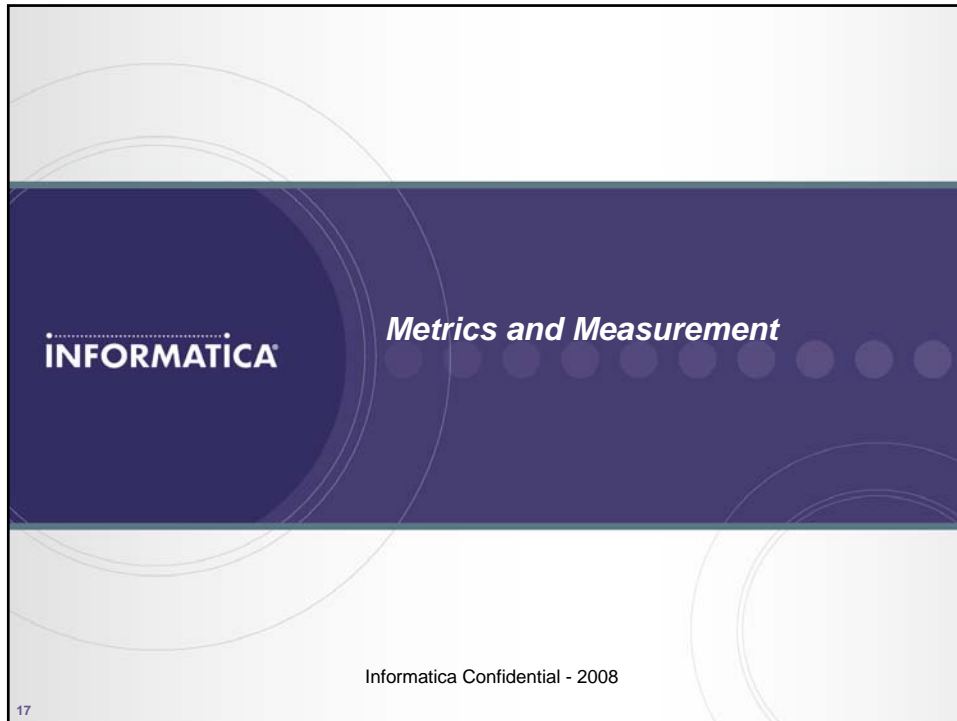


## Agenda

- Introduction
- A Pragmatic Approach
- Common Pitfalls and Misconceptions
- **Examples / Scenarios & Case Studies**
- Q&A

**INFORMATICA** Informatica Confidential - 2008 16



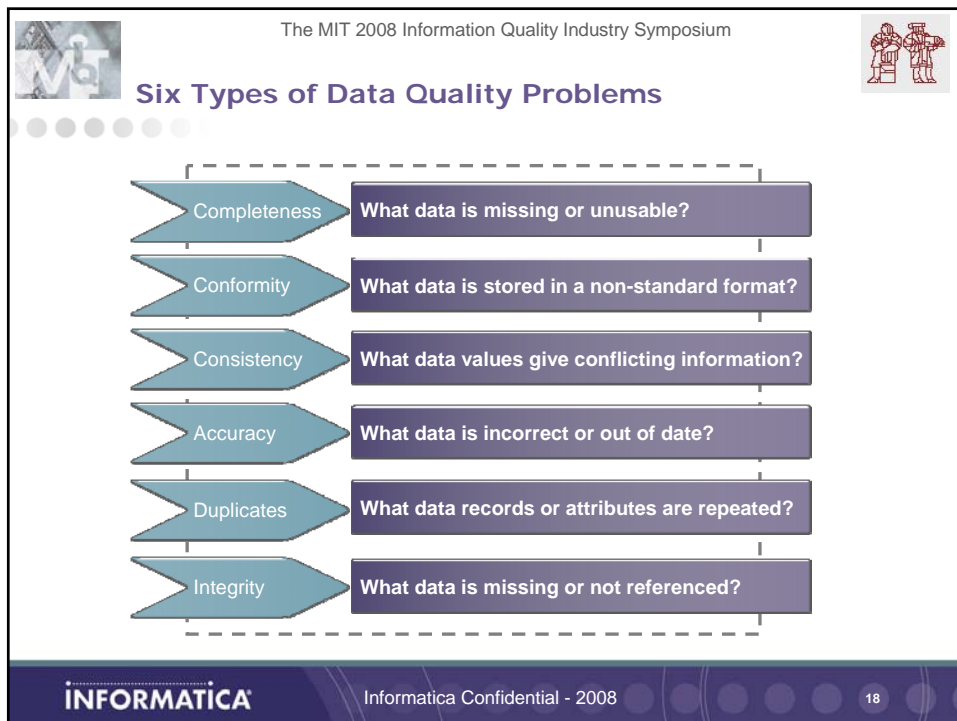


**INFORMATICA** *Metrics and Measurement*

Informatica Confidential - 2008

17

This slide features a dark blue background with the Informatica logo on the left and the title "Metrics and Measurement" in a white serif font. The background is decorated with faint, overlapping circular patterns.



The MIT 2008 Information Quality Industry Symposium

### Six Types of Data Quality Problems

Completeness	What data is missing or unusable?
Conformity	What data is stored in a non-standard format?
Consistency	What data values give conflicting information?
Accuracy	What data is incorrect or out of date?
Duplicates	What data records or attributes are repeated?
Integrity	What data is missing or not referenced?

**INFORMATICA** Informatica Confidential - 2008 18

This slide is titled "Six Types of Data Quality Problems" and lists six categories of data quality issues. Each category is represented by a teal arrow pointing to a corresponding question in a dark blue box. The categories are: Completeness (What data is missing or unusable?), Conformity (What data is stored in a non-standard format?), Consistency (What data values give conflicting information?), Accuracy (What data is incorrect or out of date?), Duplicates (What data records or attributes are repeated?), and Integrity (What data is missing or not referenced?). The slide includes the MIT logo in the top left, a small red icon in the top right, and the Informatica logo and page number "18" at the bottom.

The MIT 2008 Information Quality Industry Symposium

## Data Quality Certification Scorecard Attributes Within Object with context

Products  
Customer  
External Data  
Credit Rating

Scorecard

Completeness  
Conformity  
Consistency

Bad Record Reports  
Issue Reports

**INFORMATICA** Informatica Confidential - 2008 19

The MIT 2008 Information Quality Industry Symposium

## Data Quality Certification - Scorecard per Object



Customer Service  
AML  
Reconciliation  
Credit Risk  
Suppression List

Customer Data Scorecard

Intranet Monitor

**INFORMATICA** Informatica Confidential - 2008 20

The MIT 2008 Information Quality Industry Symposium





## Example – CPG - What is Data Quality?

Completeness	Required values electronically recorded
Standards Based	Data conforms to industry standards
Consistency	Data values aligned across systems
Accuracy	Data values are right, at the right time
Time Stamped	Validity timeframe of data is clear


*Data Quality*

**INFORMATICA** Informatica Confidential - 2008 21

The MIT 2008 Information Quality Industry Symposium



## Customer Example: AML



### CIP/KYC- Reporting Definitions

- **Completeness** - A measure of the degree to which the requested information field contains data and is non-blank.
- **Logical** - A measure of the degree to which the requested information field contains data that has been determined to be sound and reasonable within parameters defined by the responsible AML Process team.
- **Verification** – A measure of the degree to which the requested information field contains data that agrees with an appropriate authoritative source as determined by the responsible AML Process Team.
- **Accuracy** – A measure of the degree to which the requested information field contains data that agrees with the source or is consistent with fact or reality.

**INFORMATICA** Informatica Confidential - 2008 22

The MIT 2008 Information Quality Industry Symposium

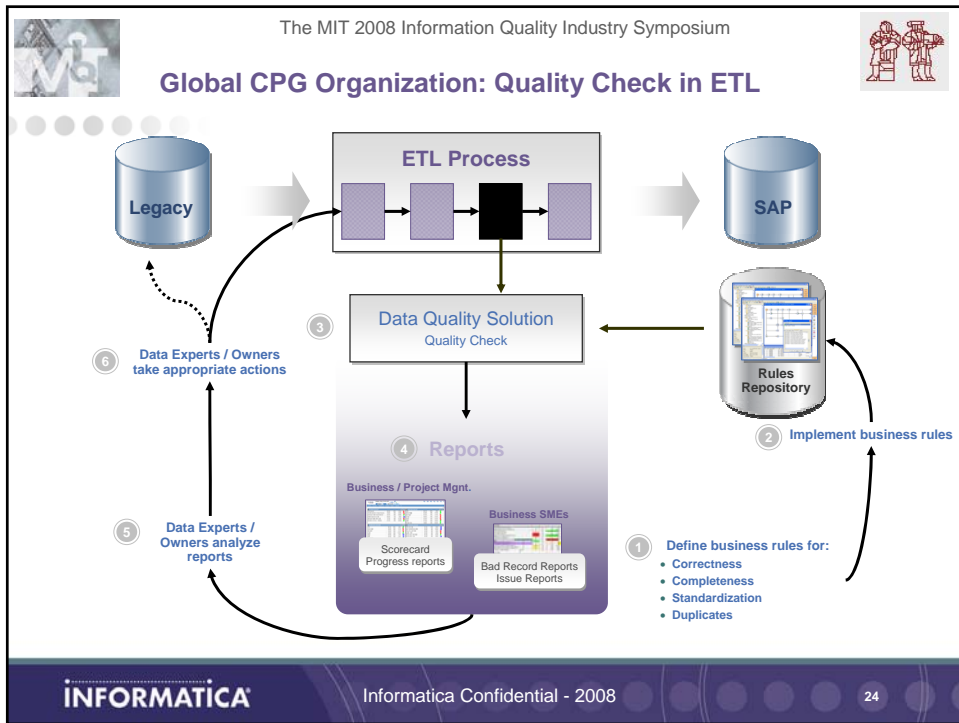
### Scorecard: Using 'Business' terms

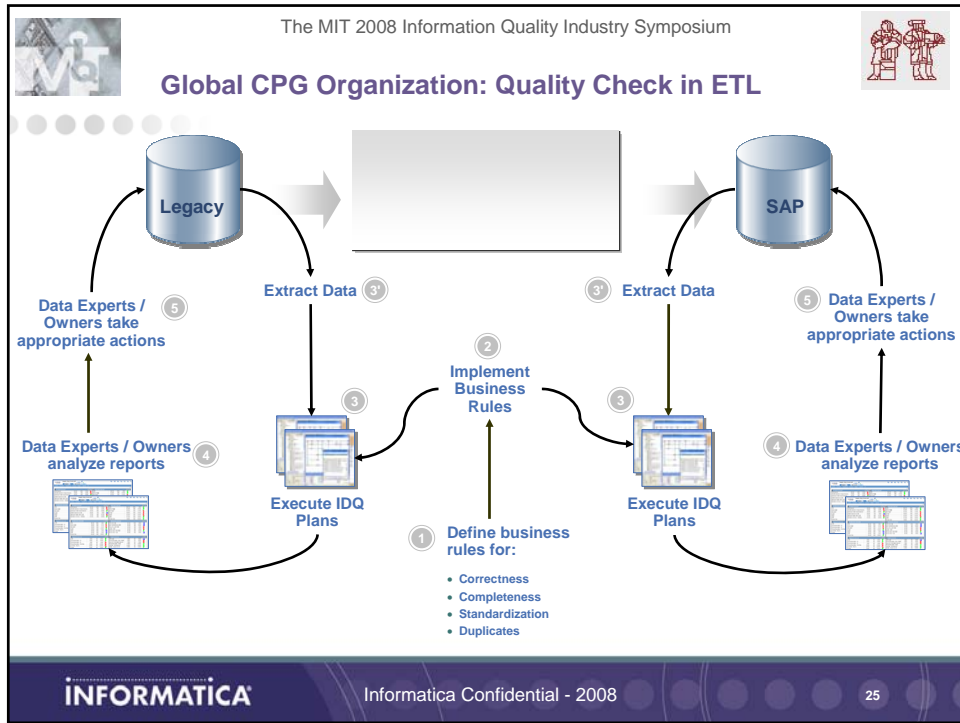
**Supply Chain Scorecard**

Conformity | Consistency | Completeness | Accuracy | Integrity | Duplicates | All



Category	Metric	Current	Target	Delta	Trend
Accuracy scorecard	GTIN Hierarchy	86.50%	100%	13.50%	Red Arrow
	All UoMs present for Product Hierarchy	45.00%	60%	15.00%	Red Arrow
	Dimensions within Product Hierarchy	48.32%	65%	16.68%	Red Arrow
	Weight Tolerances by UoM	68.00%	70%	2.00%	Blue Arrow
	BaseUoM Weight within AUoM	75.78%	90%	14.22%	Red Arrow
	Description Accuracy v. Attributes	67.40%	90%	22.60%	Green Arrow
Completeness scorecard	GTIN	98.42%	100%	1.58%	Green Arrow
	Gross Weight	75.83%	80%	4.17%	Red Arrow
	Net Weight	75.83%	80%	4.17%	Red Arrow
	Length	75.85%	90%	14.35%	Red Arrow
	Height	87.84%	90%	2.16%	Blue Arrow
	Width	87.84%	90%	2.16%	Blue Arrow
Consistency scorecard	GTIN Check Digit	93.32%	100%	6.68%	Blue Arrow
	Last Transaction Dte /Status	95.20%	80%	-15.20%	Blue Arrow
	Gross Wt => Net Weight	83.50%	80%	-3.50%	Red Arrow
	Length <= Width	85.00%	90%	5.00%	Red Arrow
	Product Class / Unit Weight	98.20%	90%	-8.20%	Green Arrow
	Product Class / Type	98.40%	90%	-8.40%	Blue Arrow
Integrity scorecard	GTIN AUoM	3.00%	0%	-3.00%	Blue Arrow
	Product Description - Dims /Weight	8.00%	0%	-8.00%	Red Arrow
	Gross Wt / Net Weight Tolerance	7.00%	0%	-7.00%	Red Arrow
	Product Description - Brand	13.00%	0%	-13.00%	Red Arrow
	Customer - D&H Hierarchy	22.00%	0%	-22.00%	Red Arrow
	Vendor - Source Cty / DC	12.00%	0%	-12.00%	Red Arrow

INFORMATICA Informatica Confidential - 2008 23





The MIT 2008 Information Quality Industry Symposium



## DQ underpins Data Governance



What we are doing at [A.N.Other CPG Org.](#)

**Achieving Data Governance thru the monitoring of the critical business rules.**

- Defining business rules to support data quality per business object (Material, Customer, Vendor, Employee, Banks etc.)
- Defining and implementing Global Data Quality and Data Management KPIs and exception reports based on the pre-Defined business rules.
- Providing Global visibility of Data Quality to the whole organisation at all levels:
  - Publishing the Data KPIs monthly on the operation site on the intranet accessible by [A.N.Other](#) organisation at all levels.
  - Presenting the Data KPIs at the Operational Steering Committee (Top Management Level) and at the market management level each month.
- Monitoring of the KPIs and the exception reports by the markets to take corrective actions.

**INFORMATICA** Informatica Confidential - 2008 27

The MIT 2008 Information Quality Industry Symposium



## Agenda

- Introduction
- A Pragmatic Approach
- Common Pitfalls and Misconceptions
- Examples / Scenarios & Case Studies
- Q&A

**INFORMATICA** Informatica Confidential - 2008 28

**INFORMATICA**

*Questions ?*

Thank You

Informatica Confidential - 2008

29