



Data Quality Management and Financial Services

Loretta O'Connor
Data Quality Sales Manager
Data Quality Division
May 2007

Content

- Introduction
- Defining the Data Quality Problem
- Solutions for Data Quality Issues
- Data Quality Reporting – Dashboards
- Data Quality Methodology – Successful
Implementing a Data Quality Strategy
- Customer Examples
- Demo
- Q&A



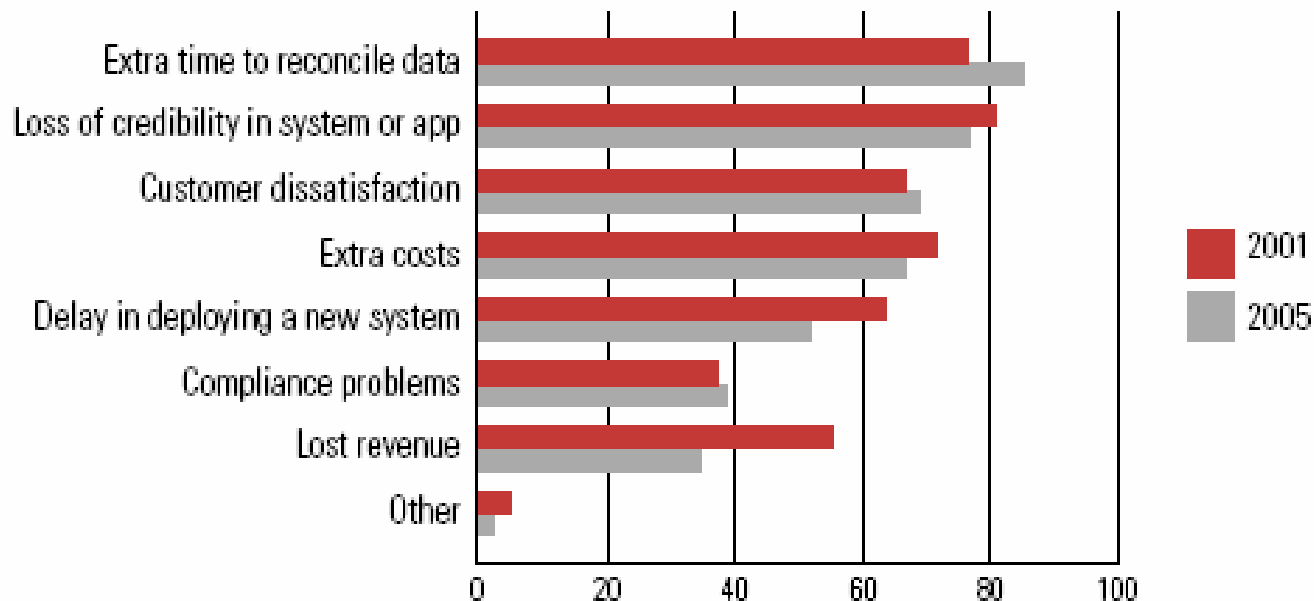


Data Quality: Problem Definition

Problem statement: Poor Data Quality causes numerous business problems



Which problems has your company suffered from due to poor-quality data?



TDWI 2006

Initiatives Driving Data Quality

Data Quality

Industry / Business Driver

- CDI, Master Data Management (All)
- Radio Frequency Identification (Manufacturing, CPG)
- Risk Management (Financial)
- Electronic availability of all services (Government)

Regulatory Compliance

- Basel II
- Sarbanes Oxley (SOX)
- Anti-Money Laundering (AML)

Internal Drivers

- Data Warehouse / BI
- Data Migrations - Mergers and Acquisitions
- Application Consolidation

The Impact

Problems

- Applications crash
- Angry business people call the operations team
- Ops track down the problems
- Problems with the accuracy of the information being reported
- Fixes being made without audit

Impact

- Applications unavailable
- Time consuming to trace and fix
- Unhappy business people
- Incorrect results
- Risk concerns
- Regulatory concerns

Root Causes

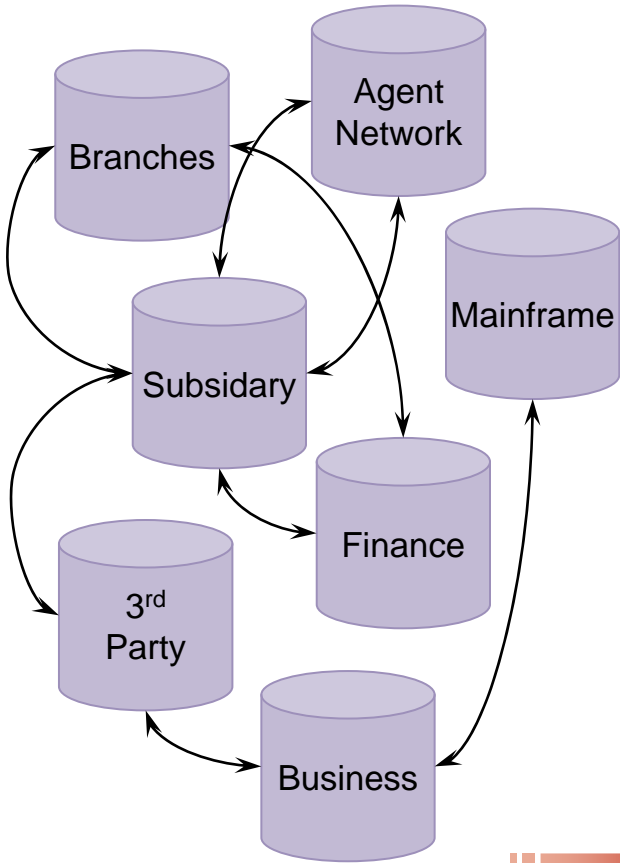
- Data didn't arrive
- Data entry errors
- Loose rules on source systems
- Data consistency errors
- File column changes
- Corrupted data

Contributory Factors

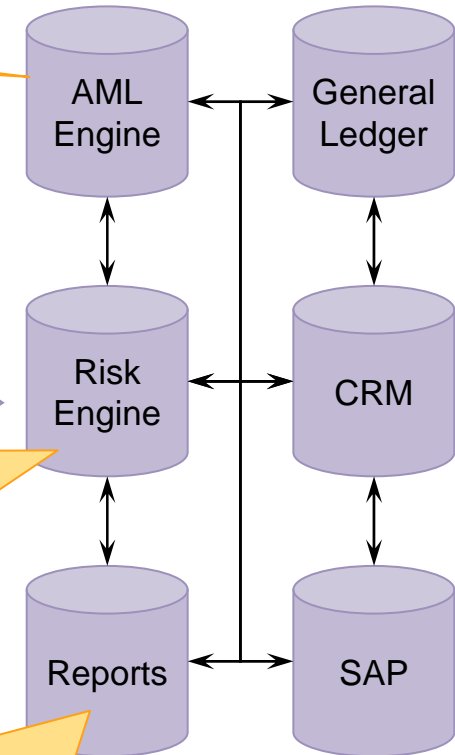
- Unclear / fragmented process
- Problem / data ownership
 - Risk operations
 - Data providers
- Multitudes of Log files

The Vicious Money Circle

Source Systems



Target Applications



“Load takes too long and this is increasing our exposure”

Analyze + Rework

Load Data

“Can't trust the data so we must manually check it”

This vicious cycle must be anticipated

“We are non-compliant and we know the regulator will see this”



Data Quality: The Solution

Existing fixes

Financial Institutions develop entire ecosystems to compensate for poor data quality

- **IT Operations**

- Unix Scripts
- Application monitoring
- Log file analysis
- Manual updates to files to 'make it work'

- All Ad Hoc
- All Manual
- Expensive to Manage
- Unreliable

- **Business**

- MS Access checks – run by business
- Manual updates to files to 'make it work'
- Same changes, every week!

And management wonder why the annual IT budget keeps getting bigger?

DQM Approach & Methodology

ANALYZE

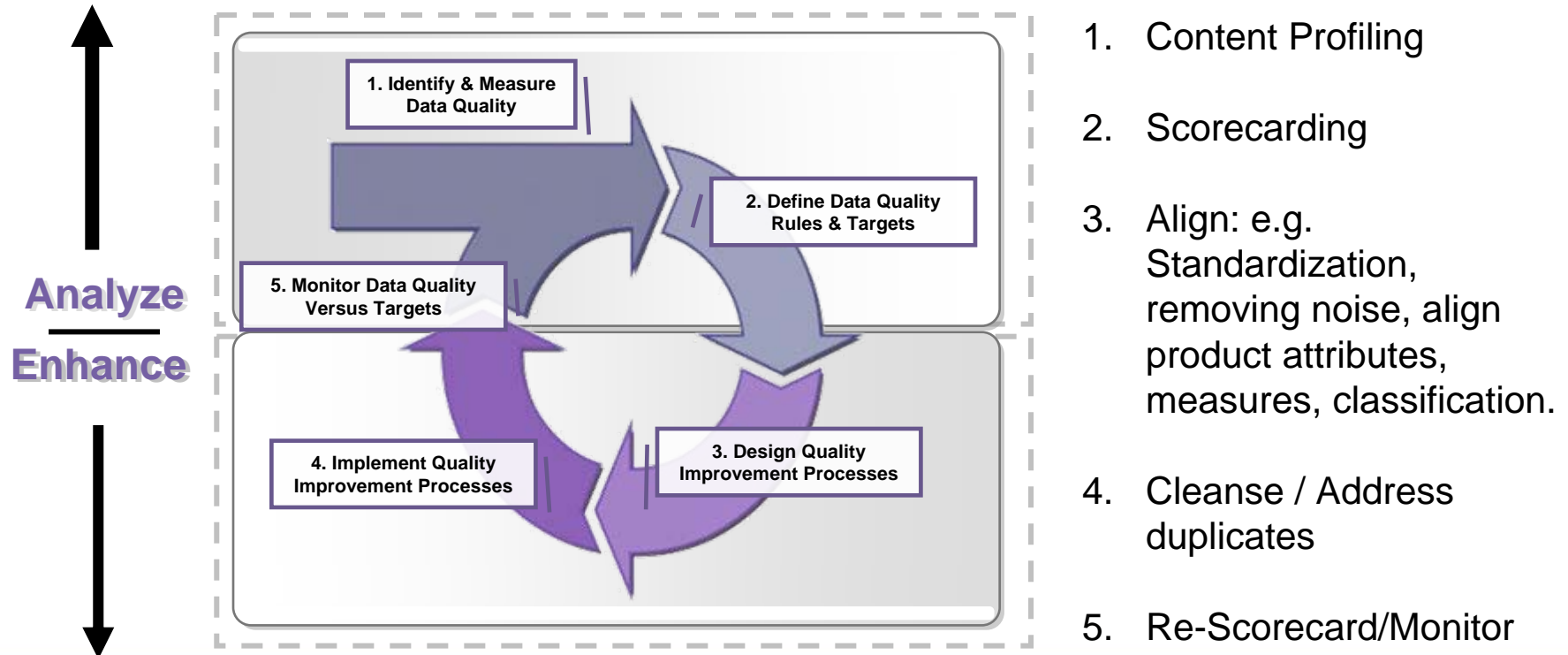
ALIGN

CLEANSE

SUSTAIN

Data Quality is not a one off exercise!

Organizations must not only align and cleanse data,
but **MUST** also keep data clean over time



Data Quality Dimensions

Data Exploration

Column Profiling

What is the data's physical characteristics ? Across multiple tables?

Relationship

What relationships exist in the data set? Across multiple tables?

Redundancy

What data is redundant? Orphan Analysis

Completeness

What data is missing or unusable?

Conformity

What data is stored in a non-standard format?

Consistency

What data gives conflicting information?

Accuracy

What data is incorrect or out of date?

Duplication

What data records are duplicated?

Integrity

What data is missing important relationship linkages?

Range

What scores, values, calculations are outside of range?

Data Quality

Sample DQ Issues

cust_no	Type	Firstname	Lastname	add1	add2	add3	zip	DOB	cust_ID	start_date	renewal_date	value	i-t-v	Rating	Alarm
15987849	Pers	ADINA	PRITCHARD	2 EWHURST AVENUE	BIRMINGHAM	Worcs	B29 6EY	05/29/1978	1	09/11/1987	04/24/2007	350,000.00	80%	A	Local
15954217	Pers	IZAAK	YAMASAKI	1 BAKER AVENUE	WHITE PLAINS							150,000.00	75%	B	None
15954218	Pers	FREDIE	MACAULAY	1455 FIRST AVENUE 2N	SAN DIEGO							1,000.00	135%	B	Monitored
15954219	Pers	DEBRAH	FIGURES	101 PARK AVENUE, SUITE	OKLAHOMA							1,000.00	200%	A	Local
15954220	Pers	LENO	TOFT	245 PARK AVE	NEW YORK							1,000.00	1%	B	None
15954221	Pers	RINALDO	HELBLING		WASHINGTON							1,000.00	null	B	Monitored
15954222	Pers				MIAMI							1,000.00	95%	D	Local
15954223	Pers				BLOOMFIELD							1,000.00	100%	B	None
15954224	Pers				PLYMOUTH							327,000.00	105%	A	Monitored
15954225	Pers				CINCINNATI			11/900	1		12/27/2007	785,000.00	65%	C	Local
15954226	Pers				PLANO		75079	1965	1		12/06/2006	470,000.00	10%	B	Local
15954227	Pers				CHICAGO	ILLINOIS	60606	1946	1	03/11/1988	07/15/2007	150,000.00	105%	B	Local
15954228	Pers				NEW YORK	NY	10038	02/12/1947	1	03/11/1988		532,000.00	10%	B	Local
15954229	Pers				NEW YORK	NY	10016	05/23/1947	1	03/11/1988		1,000.00	105%	B	None
15954230	Pers				SAN FRANCISCO	CA	94107	08/01/1947	1	03/11/1988		1,000.00	10%	A	Monitored
15954231	Pers				PHOENIX	AZ	85004	11/23/1947	1	03/11/1988		1,000.00	125%	B	Local
15954232	Pers	MARY	DAKMAN	NORTH WOODS	NEW YORK	NY	48304	09/16/1947	1	03/11/1988		1,000.00	90%	D	Local
15954233	Pers	ROMAIN	RADEL	950 3RD AVENUE	NEW YORK	NY	10022	11/01/1947	1	03/11/1988		1,000.00	90%	B	Monitored
15954234	Pers	MCKINLEY	OKUMURA	11 WALL STREET 11TH	NEW YORK	NY	10004	02/13/1945	1	03/11/1988		785,000.00	180%	B	Local
15954235	Pers	THORNDIKE	BRANNAN	990 STEWART AVENUE	GARDEN CITY	NY	11530	07/07/1966	1	07/1981	05/21/2007	470,000.00	75%	B	Local
15954257	Pers	KY	SCHNEIDER	1037 PARKVIEW DRIVE	COVINGTON	LA	70424		1	04/13/2007		610,000.00	110%	A	None
15954258	Pers	SHEPLEY	HINZE	29 WINFIELD AVENUE	HARTFORD	CT	06112		1	08/11/2007		532,000.00	70%	B	Local
15954259	Pers	CAROLE	TWITCHELL	270 PARK AVENUE	HARTFORD	CT	06112		1	11/2007		500,000.00	55%	A	Local
15954260	Pers	TED	MC AUCLAY	41 LEXINGTON AVE.	NEW YORK	NY	10017		1	2/06/1985		278,000.00	95%	B	None
15954261	Bus	MARILOU	MARGHERIO	1655 LA FONDA	LOS ANGELES	CA	90011		1	07/15/2007		453,000.00	100%	B	Monitored
15954262	Pers	LATI	SCOTTO	75 MAIN STREET	COVINGTON	LA	70424		1	06/11/2007		1,000.00		D	Local
15954263	Pers	NEILS	TEASTER	220 EAST 42ND STREET	NEW YORK	NY	10017	12/12/1967	1	05/08/1985	04/24/2007	785,000.00	115%	B	None
15954264	Pers	OZA	PETRUCCI	3250 WILSHIRE BLVD.	LOS ANGELES	CA	90010-1438	03/19/1965	1	13/08/1985	10/19/2006	470,000.00	90%	B	Monitored
15954265	Pers	DAMALI	ROOM	12140 ARTESIA BLVD., STE 107	ARTESIA	CA	90701	07/11/1946	1	18/08/1986	09/23/2007	610,000.00	80%	C	Local
15954266	Pers	HOWLAN	DRUCKER		RANDOLPH	VT	05035	02/12/1947	1	02/12/1986	03/11/1988	532,000.00	175%	B	None
15954267	Pers	KINNIE	CYPHERS	909 W. 9TH STREET	ANCHORAGE	AK	99501	03/19/1965	1	03/12/1987	03/29/1988	250,000.00	70%	B	Local
15954268	Pers/Bus	RUFENA	HUFFORD	51 CANDLEWOOD DRIVE	BARRINGTON	IL	60010	01/01/1900	1	05/01/1987	02/05/2007	276,000.00	85%	B	Local
15954269	Pers	LORE	VIRDEN	88 SHREVEPORT RD.	BARKSDALE	LA	71110-2090	02/12/1947	1	14/01/1987	05/21/2007	1,300,000.00	60%	A	None
15954270	Pers	BELVA	STEINERT	1111 STEWART AVENUE	BETHPAGE	NY	11714-3581	05/29/1978	1	20/11/1988	04/13/2007	900,000.00	95%	B	Monitored
15954271	Pers	ADAMSEN	REISINGER	82 DEVONSHIRE STREET E31A	BOSTON	MA	02109	08/01/1959	1	31/08/1989	08/11/2007	500,000.00	105%	B	Local

Duplication:
Fuzzy matching

Conformity:
Incorrect Format

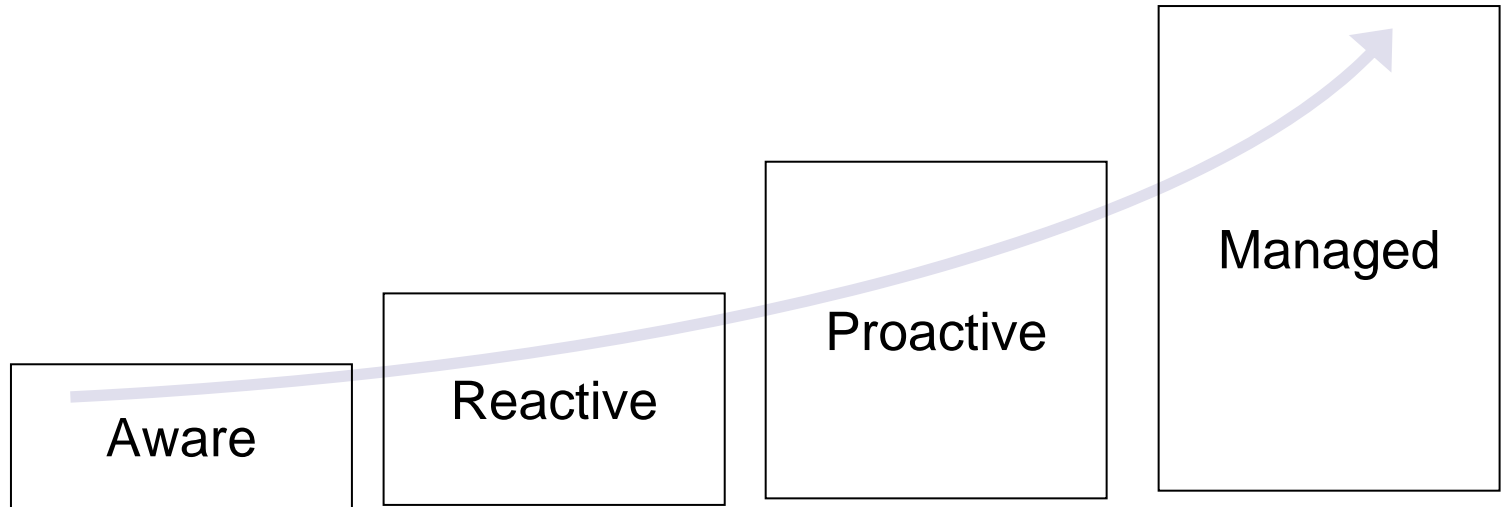
Consistency:
Data is in correct format and complete, but breaks a business rule

Range:
Identify outliers

Accuracy:
Using reference data to validate

- COMPLETENESS
- CONFORMITY
- CONSISTENCY
- DUPLICATION
- INTEGRITY
- ACCURACY
- RANGE

Data Quality Maturity Model



Attitude	• DQ seen as a cost	• DQ for IT	• DQ driven by business	• DI and DQ seen as key enabler
Tech.	• Hand coded	• Silos of DQ	• Linked projects	• Fully integrated DQ initiative
Benefit	• Few	• Few tactical	• Key tactical gains	• Strategic

Drivers depend on where you are and where you want to go



Sample Financial Services Business Intelligence Dashboards

IDQ: Data Accuracy Scorecard

Conformity

Trend	Item	Passed %	Target %	40%	100%	Jun 23, 2006	Jun 20, 2006	Jun 19, 2006
▲	Weighted Average	84.3	90.0			84.3	84.3	84.3
	CL_Subscriber_SSN	96.9	90.0			96.9	96.9	96.9
	CL_Claim_SSNSubscriber_Nbr	96.9	90.0			96.9	96.9	96.9
▲	CL_Claim_Birth_Date	100.0	90.0			100.0	100.0	100.0
	CL_Claim_ICD1_Code	44.7	90.0			44.7	44.7	44.7
	CL_Claim_ICD2_Code	66.1	90.0			66.1	66.1	66.1
	CL_Claim_ICD3_Code	83.0	90.0			83.0	83.0	83.0
	TL_FirstName	87.1	90.0			87.1	87.1	87.1
	CL_Claim_Client_Name	100.0	100.0			100.0	100.0	100.0

Integrity

Trend	Item	Passed %	Target %	40%	100%	Jun 23, 2006	Jun 20, 2006
	Weighted Average	99.3	95.0			99.3	99.3
▲	RBA_Claim_Birth_Date_Validation	99.2	95.0			99.2	99.2
	Claim_Client_Name_Validation	100.0	100.0			100.0	100.0
	Claim_Client_Type_Code_Validation	99.9	100.0			99.9	99.9
▲	RBA_Claim_Source_Name_Validation	99.2	100.0			99.2	99.2
	RBA_Claim_Subscrbr_SSN_Nbr_Validation	96.9	95.0			96.9	96.9
▲	RBA_Claim_Status_Code_Validation	100.0	95.0			100.0	100.0
	RBA_Claim_Relationship_Code_Validtn	99.6	95.0			99.6	99.6
	RBA_Claim_RecordType_Code_Validtn	99.9	95.0			99.9	99.9

3rd Party Reporting using IDQ

INFORMATICA
The Data Integration Company

Basel 2 data quality | Operational risk | Alerts

Informity | Consistency | Completeness | Accuracy | Integrity | Duplicates | All

Custo

Accuracy scorecard					Conformity scorecard			
Accuracy of Credit rating		50.19	100		-49.81	Conformity of Credit rating	10.62	
Accuracy of EAD		10.3				Conformity of EAD	3.93	
Accuracy of Exposure Amount		13.13				Conformity of Exposure Amount	2.03	
Accuracy of LGD		5.59				Conformity of LGD	2.32	
Accuracy of Maturity Date		31.85				Conformity of Maturity Date	8.08	
Accuracy of PD		6.62				Conformity of PD	2.45	

Completeness scorecard					Consistency scorecard			
Completeness of Credit rating		87.27	100		-12.73	Consistency of Credit rating	126.44	
Completeness of EAD		14.12				Consistency of EAD	19.88	
Completeness of Exposure Amount		27.42				Consistency of Exposure Amount	37.84	
Completeness of LGD		18.83				Consistency of LGD	22.88	
Completeness of Maturity Date		55.33	60		-4.67	Consistency of Maturity Date	78.54	
Completeness of PD		8.2				Consistency of PD	11.98	

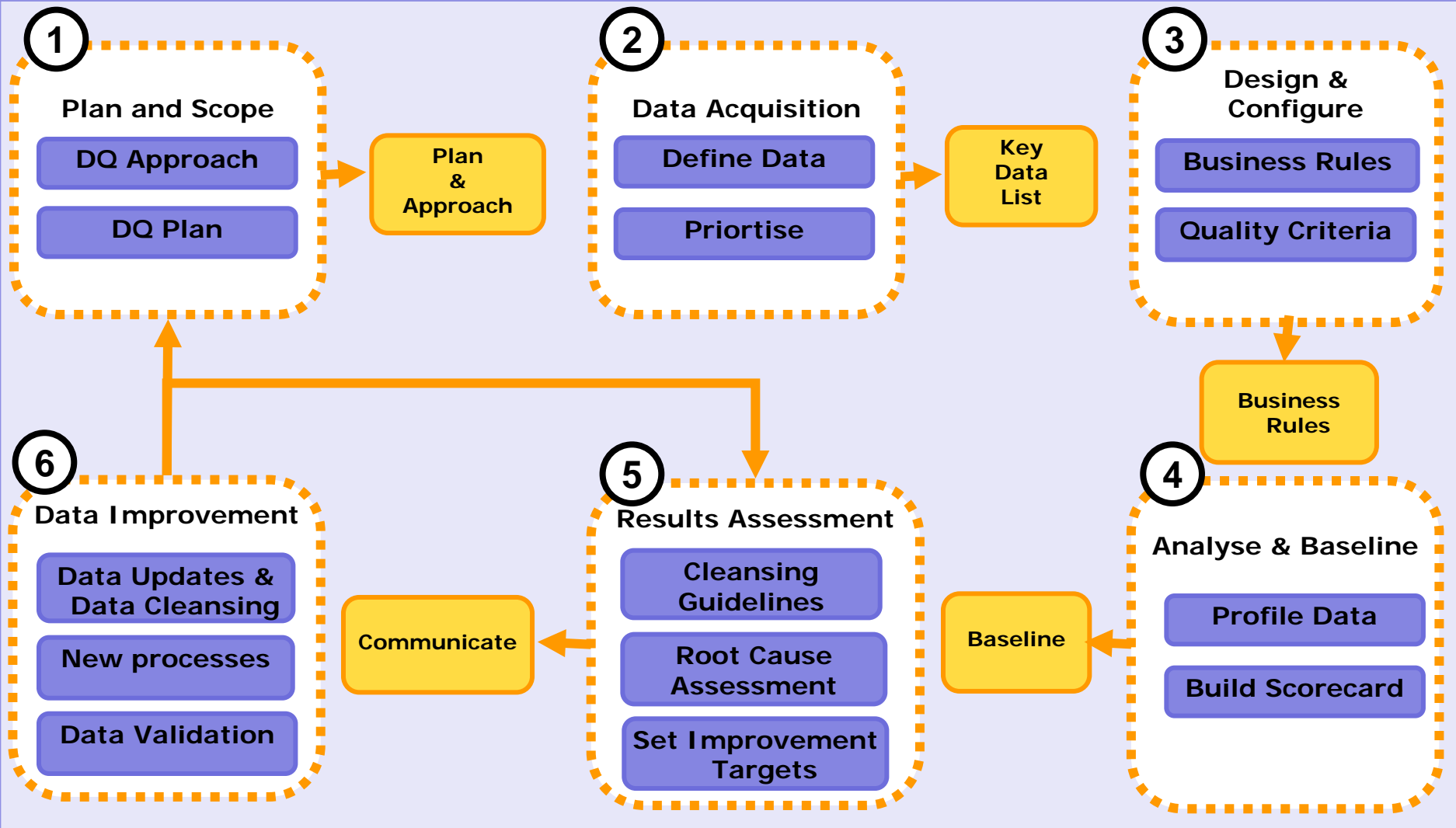
Duplicates scorecard				Integrity scorecard		
Duplicates of Credit rating		39.17		Integrity of Credit rating	42.33	
Duplicates of EAD		5.77		Integrity of EAD	11.62	
Duplicates of Exposure Amount		10.41		Integrity of Exposure Amount	45.74	
Duplicates of LGD		4.05		Integrity of LGD	10.98	
Duplicates of Maturity Date		23.21		Integrity of Maturity Date	33.26	



Methodology



Scorecarding Back to Source™





Customers



Master Data Management

Improve

- Enable business user to build data quality monitoring rules
- Provide standard platform that could be extended for further data quality initiatives

Cadbury Schweppes

Challenge

- **Problems managing trade promotions because of poor data quality**
- **Data migrations put at risk because of data quality issues**

How We Helped

- **Ability to monitor and cleanse all types of data product, customer and business**
- **Flexibility to manage and control different data quality problems on one platform**

Business Value

- **Data quality improvement leads to more streamlined supply chain**
- **Faster more successful data migrations and systems consolidation**

Third Largest Bank in the US

Informatica In Action



KEY BUSINESS IMPERATIVE

Regulatory Compliance

- Compliance with anti-money laundering regulations
- Provide robust DQ reporting and metrics system for AML Unit

IT/BUSINESS INITIATIVE:

Regulatory Reporting

DATA QUALITY INITIATIVE:

DQ Reporting & Monitoring

THE CHALLENGE

- Enable AML team to build, manage and customize AML business rules
- Track and monitor data quality across key systems

INFORMATICA ADVANTAGE

- Data quality workbench for business users
- Scorecard aggregating data quality metrics from multiple systems

RESULTS/BENEFITS

- Avoided regulatory penalties of up \$20m
- Implemented AML DQ Monitoring ahead of deadline using existing AML team resources
- Saved estimated \$3m+ cost of bespoke of AML solution

Reuters: Global CRM management



Key Business Requirements:

- “Fix data quality within existing Siebel systems”

Approach:

- Provide data quality metrics to drive improvement processes
- Implement one off and ongoing data quality processes

Challenge

- Lack of ROI on Siebel due to low quality data
 - Poor client management
 - Inaccurate mailing processes
 - Inefficient marketing processes
- The manual generation of monthly data quality reports very inefficient.

Solution

- Informatica Data Quality
 - To implement an automated Data Quality Scorecard per country
 - To implement one off and then ongoing cleansing and standardization
- Informatica Data Explorer
 - To profile new data sources

Expected Results

- Increase in sales force and marketing efficiency
- Recognised Data Quality metrics process in place