Data Quality Management and Financial Services

INFORMATICA
financial services practice

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

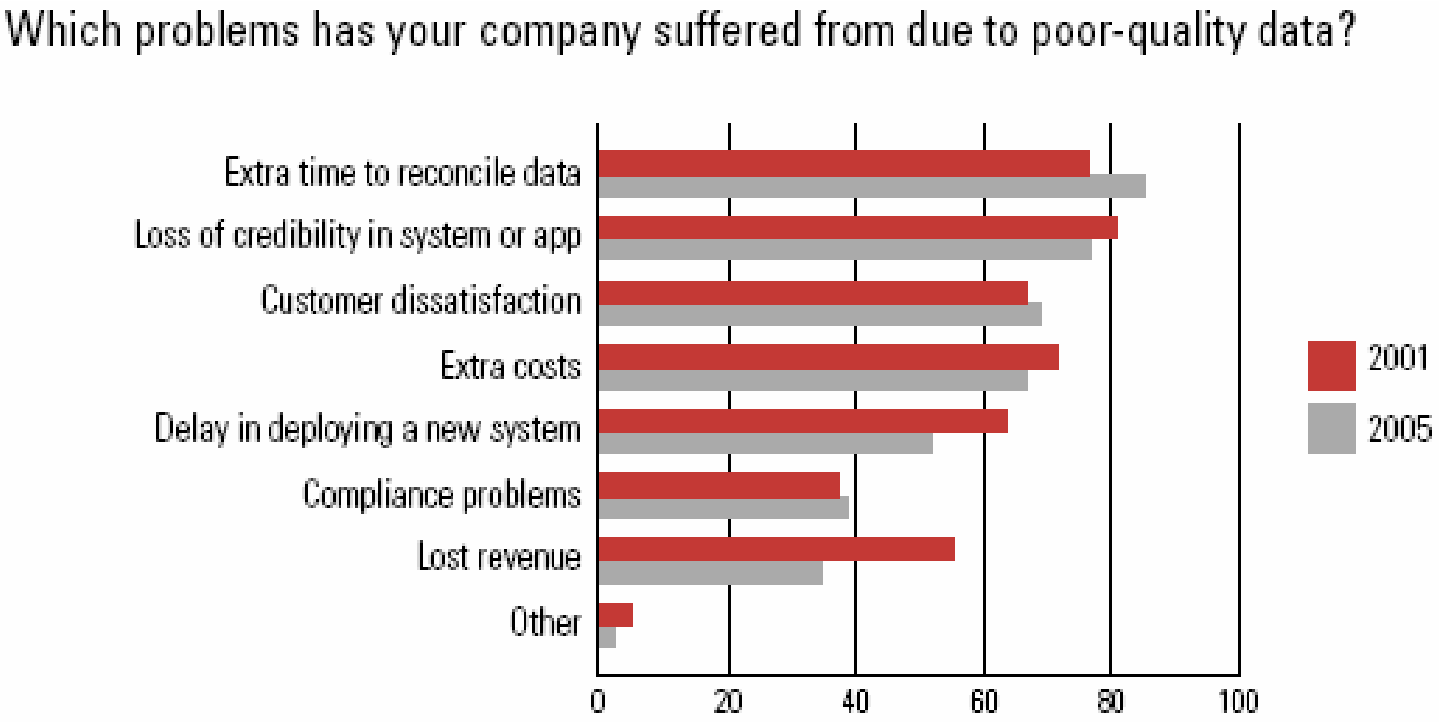
Content

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

INFORMATICA
financial services practice

Data Quality: Problem Definition

Problem statement:
Poor Data Quality causes numerous business problems



TDWI 2006

Initiatives Driving Data Quality

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

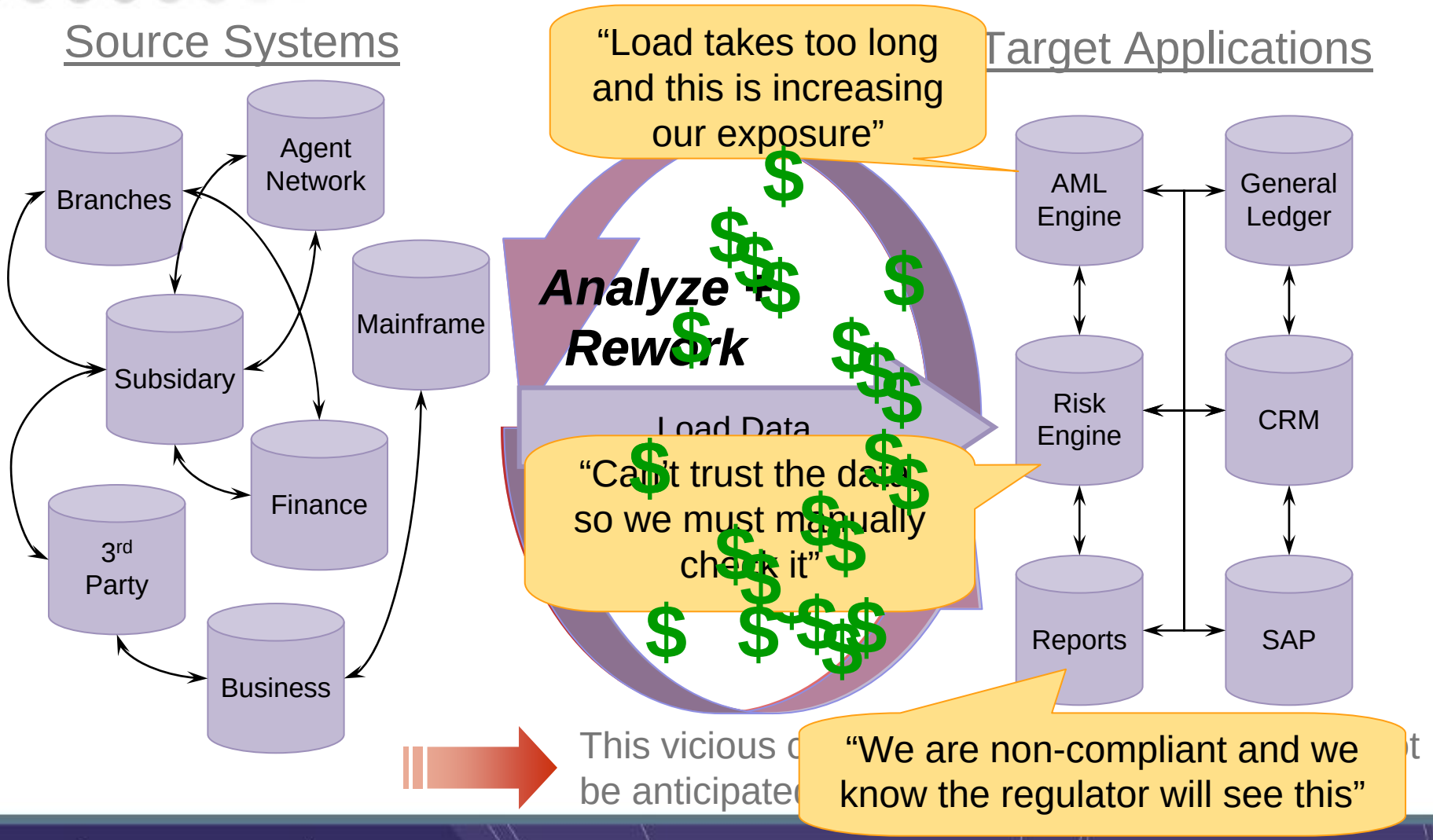
Root Causes

- 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



INFORMATICA
financial services practice

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

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

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

DQM Approach & Methodology

ANALYZE → ALIGN → CLEANSE → SUSTAIN

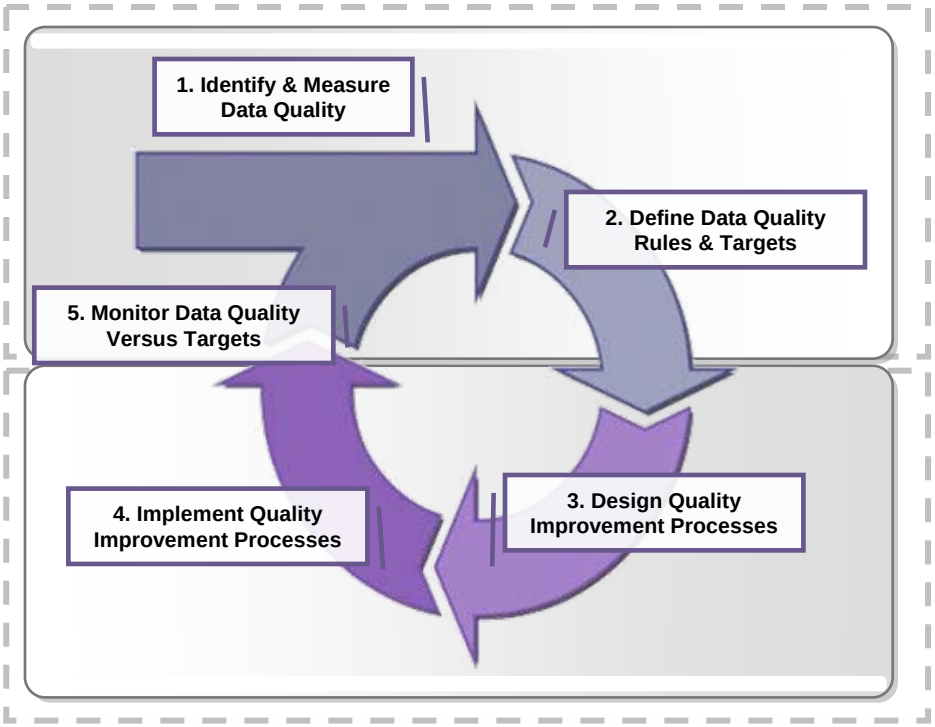
Data Quality is not a one off exercise!

Organizations must not only align and cleanse data,

but MUST also keep data clean over time

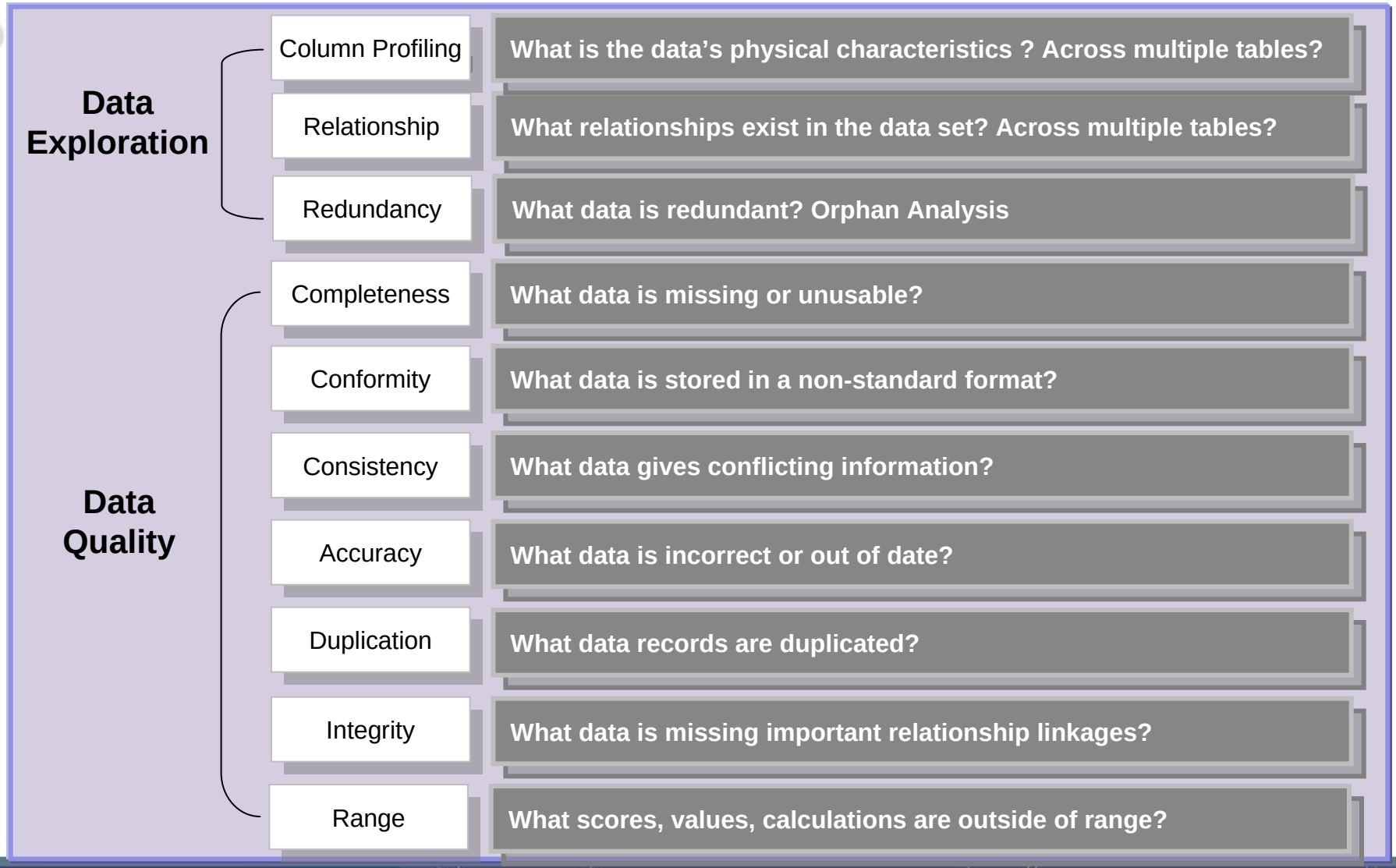
Analyze

Enhance



1. Content Profiling
2. Scorecarding
3. Align: e.g. Standardization, removing noise, align product attributes, measures, classification.
4. Cleanse / Address duplicates
5. Re-Scorecard/Monitor

Data Quality Dimensions



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							000.00	75%	B	None
15954218	Pers	FREDIE	MACAULAY	1455 FIRST AVENUE 2N	SAN DIEG							.00	135%	B	Monitored
15954219	Pers	DEBRAH	FIGURES	101 PARK AVENUE, SUITE	OKLAHOM							.00	200%	A	Local
1595422C	Pers	LENO	TOFT	245 PARK AVE	NEW YORI							.00	1%	B	None
15954221	s	RINALDO	HELBLING		WASHING							.00	null	B	itored
15954222	Pe				MIAMI							.00	95%	D	ocal
15954223	Pers				BLOOMFIE							.00	100%	B	None
15954224	Pers				PLYMOUTH							327,000.00	105%		Monitored
15954225	Pers				CINCINNATI			1900	1		27/2007	785,000.00	65%	C	Local
1595					PLANO		7507	1965	1	1	12/06/2006	470,000.00	10%	B	Local
1595					CHICAGO	ILLINOIS	60606	946	1	03	07/15/2007	,000.00	105%	B	Local
1595					NEW YORK	NY	10038	02/12/				532,000.00		B	Local
1595					NEW YORK	NY	10016	05/29				.00	105		None
1595				E	SAN FRANCISCO	CA	94107	08/01				.00	10		Monitored
1595					PHOE	AZ	85004	11/23				.00	125%	B	Local
15954232	Pers	MARY	OAKMAN	NORTH WOOD	NEW YO	NY	48304	09/16				.00	90%	D	Loc
15954233	Pers	ROMAIN	RADEL	950 3RD AVENUE	NEW YORI		1002	11/0				.00	90%	B	M ed
15954234	Pers	MCKINLEY	OKUMURA	11 WALL STREET 11T	NEW YORK			02/13/1945				785,000.00	180%	B	
	Pers	THORNDIKE	BRANNAN	990 STEWART AVENU	GARDEN CIT			07/07/1966		07/1981	05/21/2007	470,000.00	75%	B	e
15954257	Pers	KY	SCHNEIDER	1037 PARKVIEW DRIVE	CO						04/13/2007	610,000.00	110%	A	None
15954258	Pers	SHEPLEY	HINZE	29 WINFIELD AVENUE	H					08/11/	07	532,000.00	70%	B	Local
15954259	Pers	CAROLE	TWITCHELL	270 PARK AVENUE	M					01/2		500,000.00	55%	A	Local
15954260	Pers	TED	MC AUCLAY	41 LEXINGTON AVE.	M					2/06/		278,000.00	95%	B	None
15954261	Bus	MARILOU	MARGHERIO	1655 LA FONDA	L					7/15/20		453,000.00	100%	B	Monitored
15954262	Pers	LATI	SCOTTO	75 MAIN STREET	CO						06/11/2007			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/ 07	610,000.00	80%	C	Local
15954266	Pers	HOWLAN	DRUCKER		RADNOR	P	19087	02/12/1947	1	02/12/1986	03/	532,000.00	175%	B	None
	Pers	KINNIE	CYPHERS	909 W. 9TH STREET	ANCHORAGE	AK	99501	03/19/1965	1	03/12/1987	03/29/	250,000.00	70%	B	Local
15954268	Pers/Bus	RUFENA	HUFFORD	51 CANDLEWOOD DRIVE	BARRINGTON	IL	60010	01/01/1900	1	05/01/1987	02/05/00	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		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	2109	08/01/1959	1	31/08/1989	08/11/2007	500,000.00	105%	B	Local

Duplication:
Fuzzy matching

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

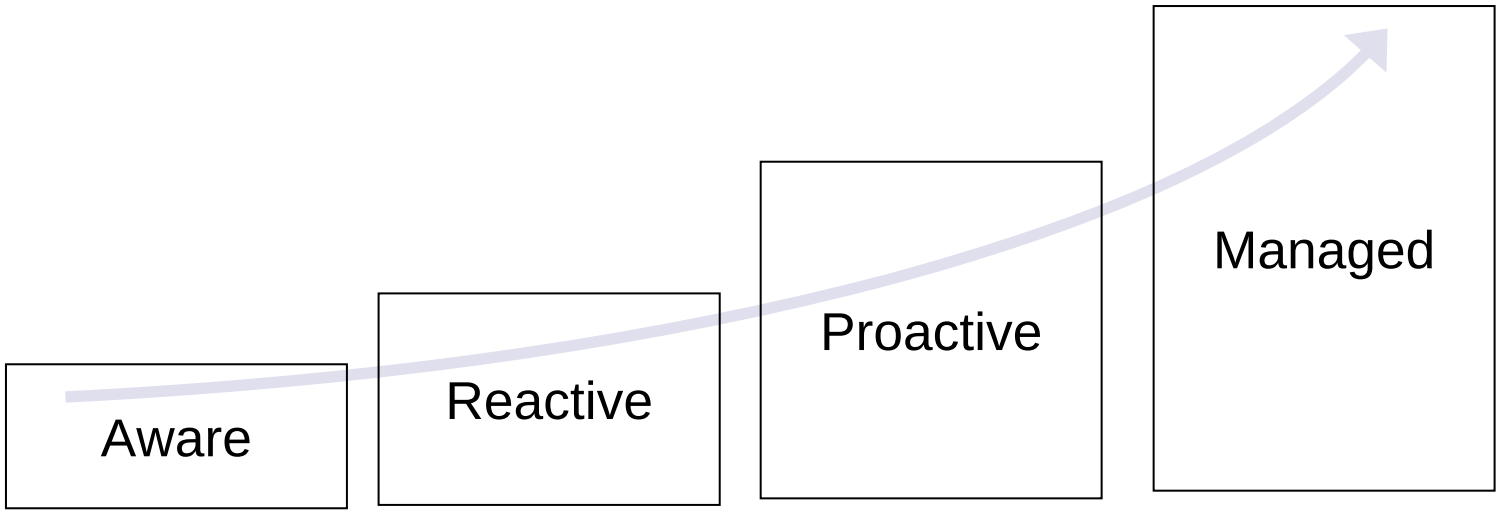
Conformity:
Incorrect Format

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

INFORMATICA
financial services practice

Sample Financial Services Business Intelligence Dashboards

IDQ: Data Accuracy Scorecard

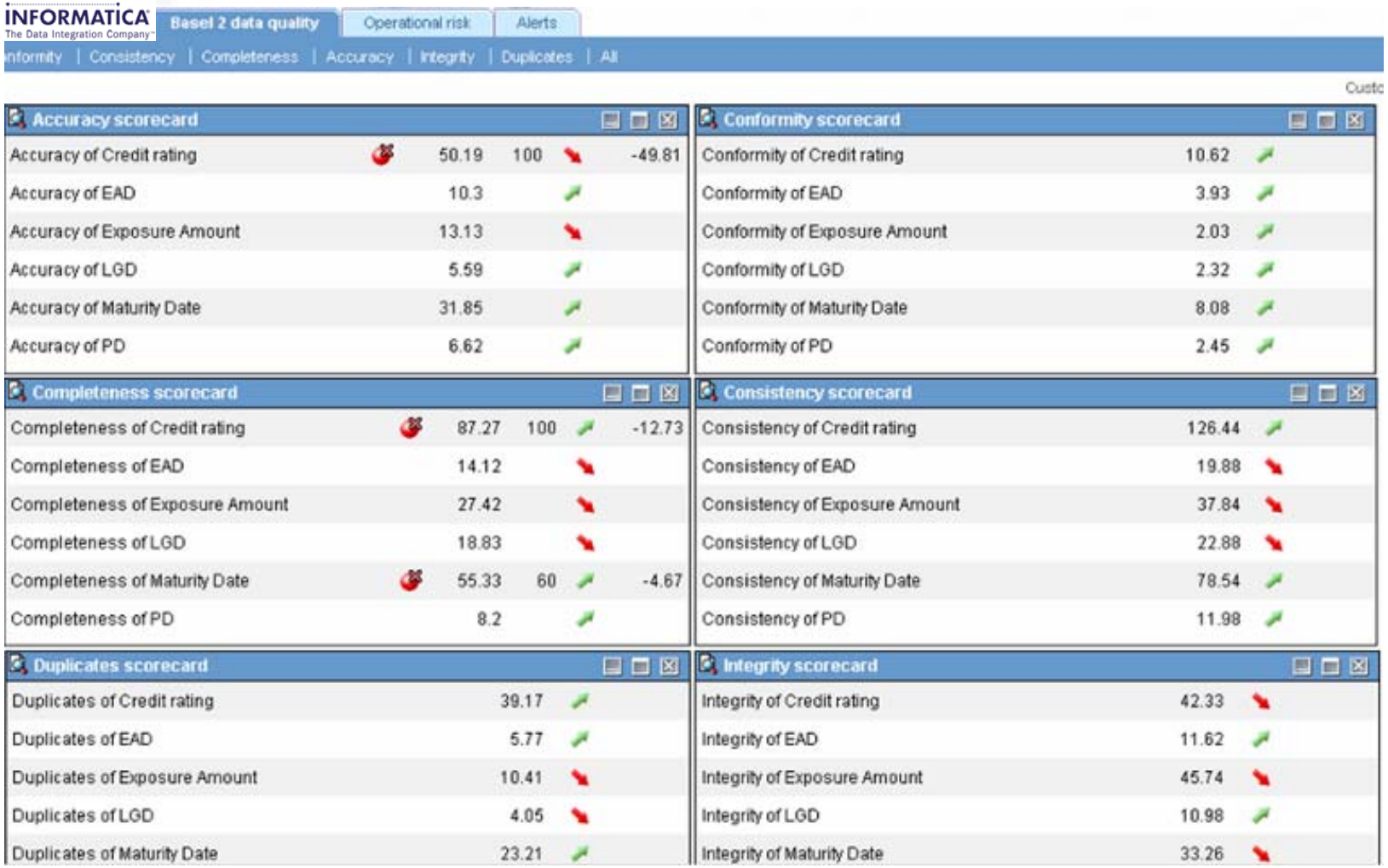
Conformity

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

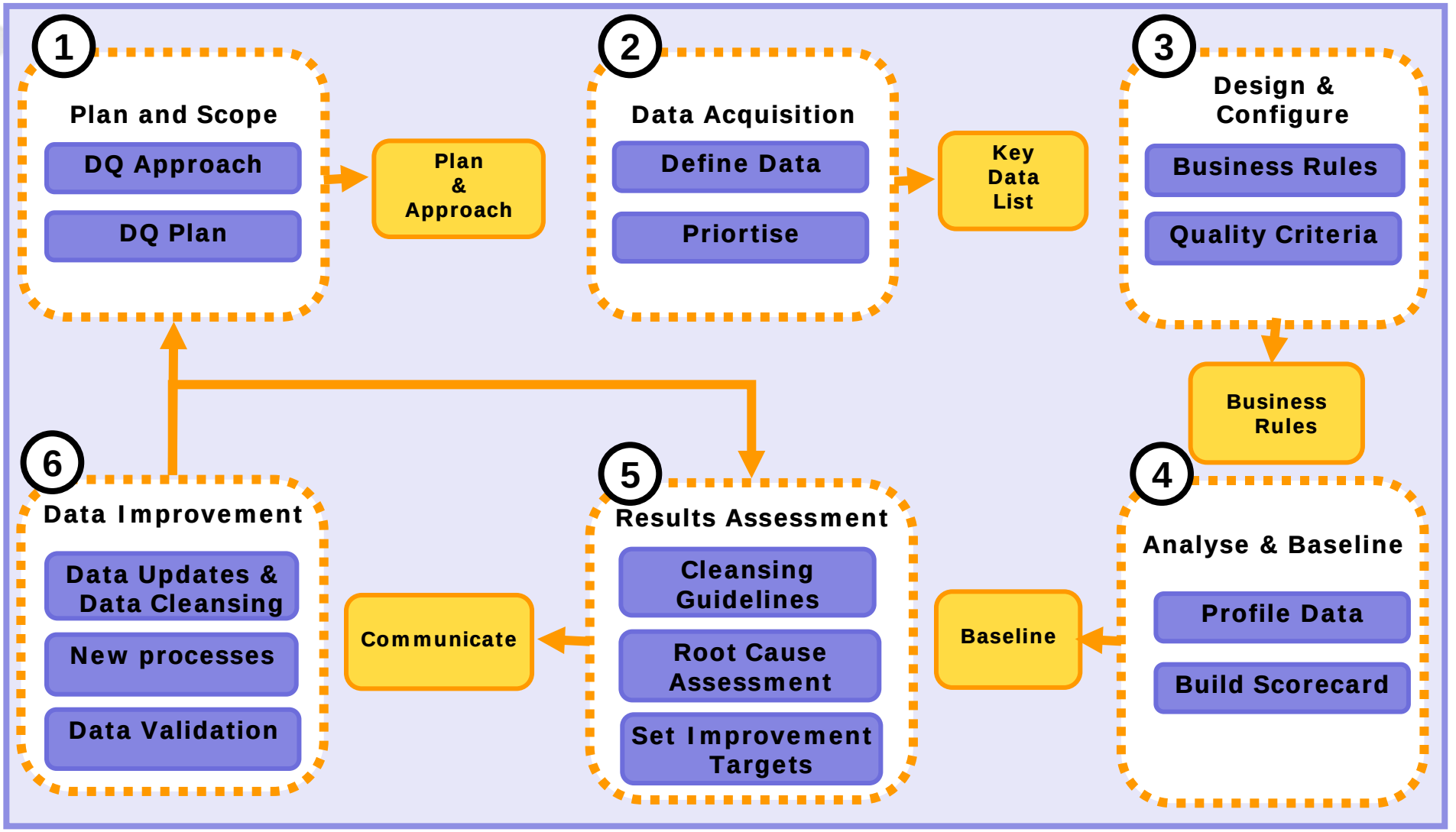
3rd Party Reporting using IDQ



Methodology

INFORMATICA
financial services practice

Scorecarding Back to Source™



Customers

INFORMATICA
financial services practice

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

IT/BUSINESS INITIATIVE:

Regulatory Reporting

DATA QUALITY INITIATIVE:

DQ Reporting & Monitoring

KEY BUSINESS IMPERATIVE

Regulatory Compliance

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

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