Data Quality, An Enterprise Principle and Practice

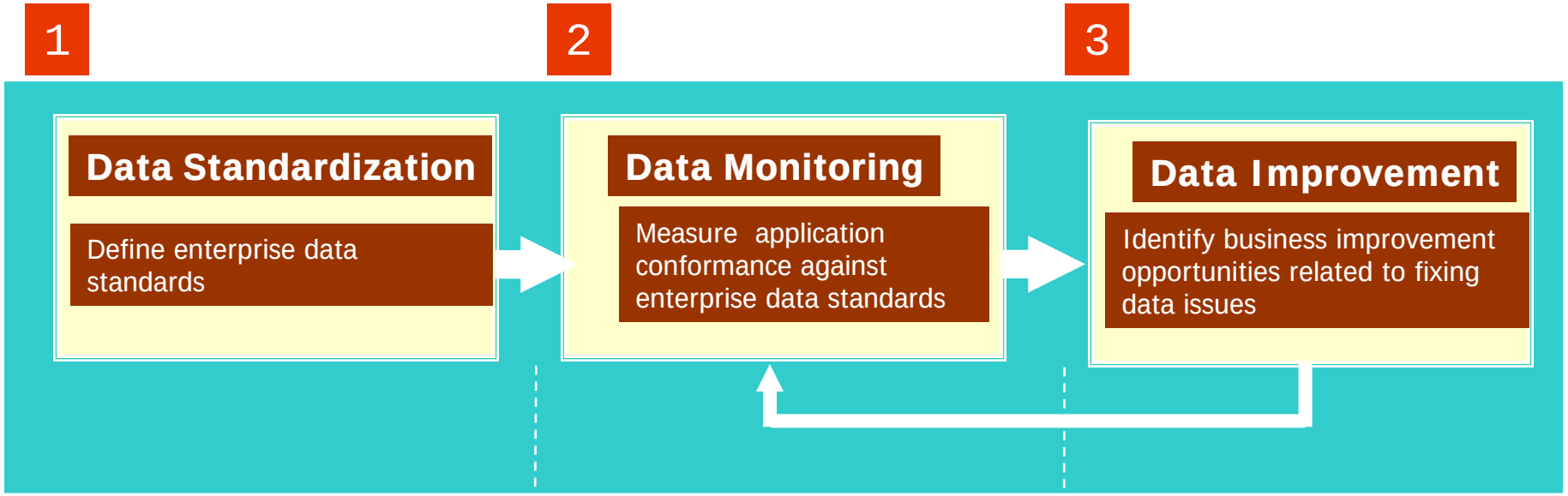
Chris Michel

July 2007

Objectives of this presentation

- **Contribute to the research on Information Quality (IQ), providing a working example of Enterprise Data Quality**
- **IQ Framework Methodology and Process**
- **IQ Roles and Responsibilities**
- **IQ Quality Scorecard**
- **Discussion**

Data Quality Framework

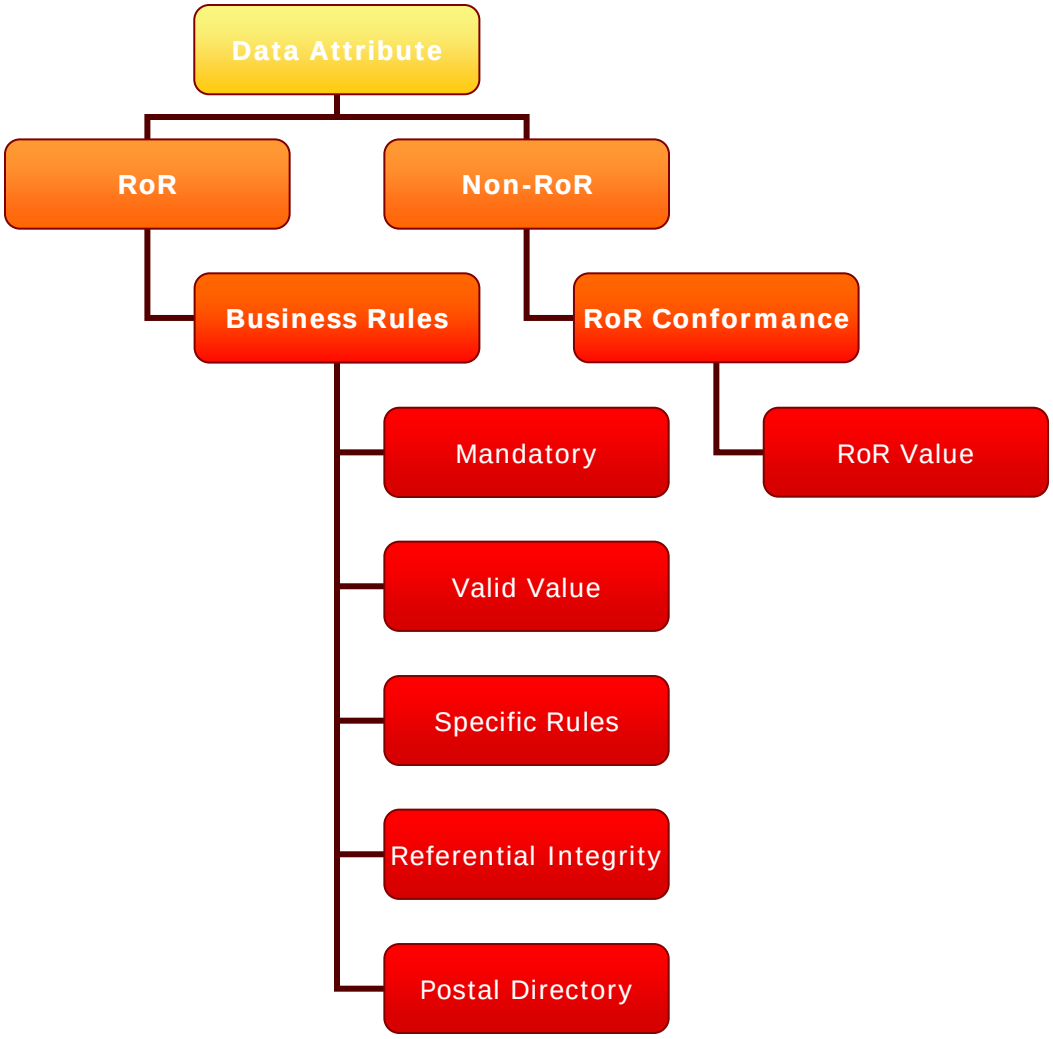


- Key Data Subjects (8)
- Key Data Attributes for each Data Subject (300+)
- Key Business Rules for each Data Attribute (1000+)

- Key enterprise applications (8)
- Audit business rules in RoR applications
- Audit RoR compliance in Non-RoR applications

- Significant project recommendations
- Minor grooms/fixes

Data Quality Metrics Methodology



Metric Components:

Business Rules

- Is the value null or blank?
- Does the value match a list of specified valid values?
- Does the value conform to specific NCR business rules?
- Does the value cause a referential integrity concern?
- Does the value conform to postal standards? (address elements only)

RoR Conformance

- Does the value in the application match the value in the RoR

Metric Calculations:

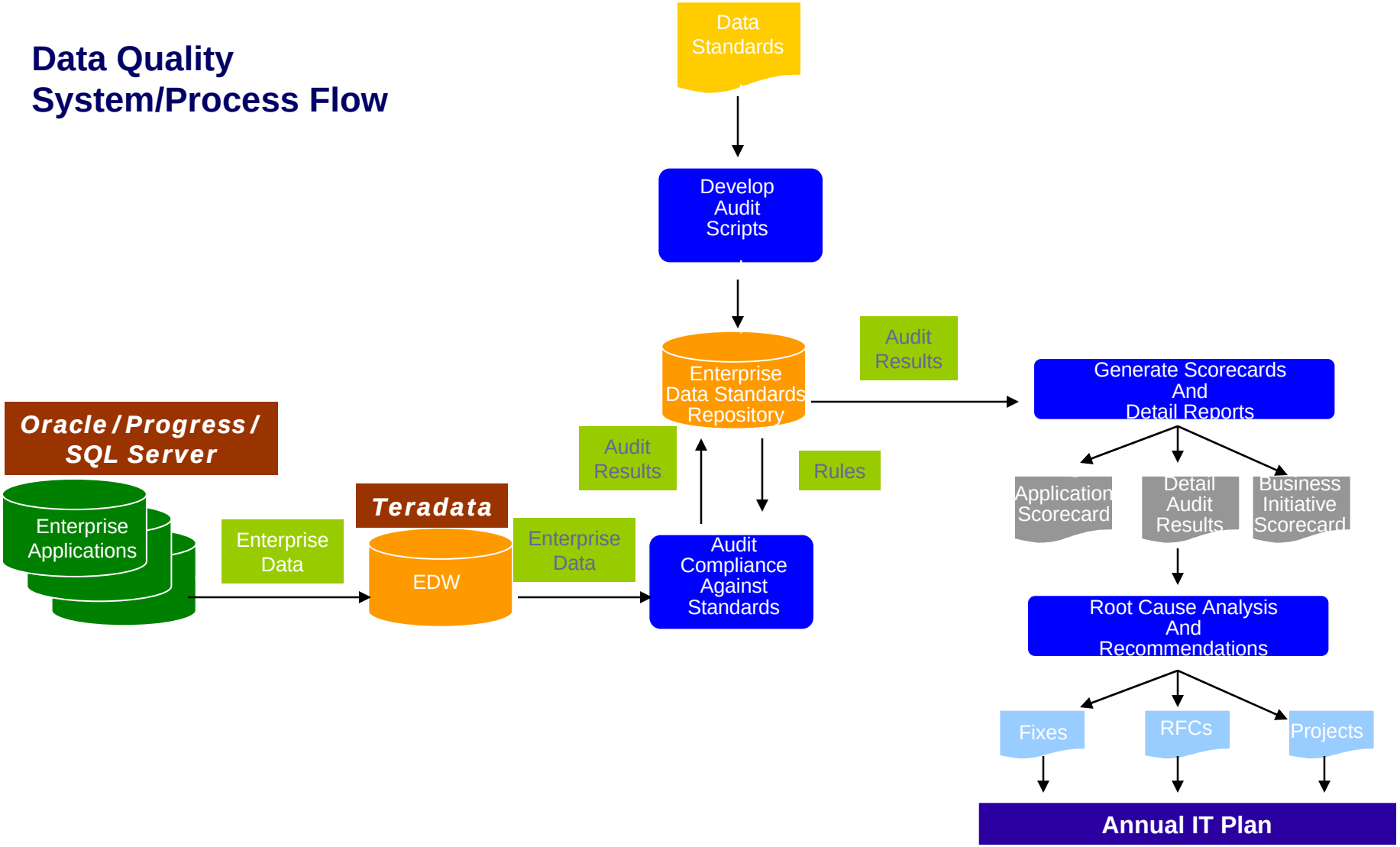
Component Level

- Number of Rows Passing Check / Number of Rows Checked

Aggregate Level

- Sum(Number of Rows Passing All Checks) / Sum(Number of All Rows Checked)

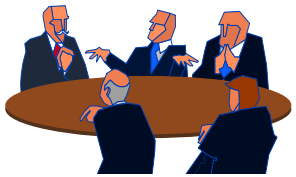
Data Quality System/Process Flow



Data Quality Roles & Responsibilities

Data Subject Experts (Business and/or ITS)

- Provides business subject matter expertise to define data standards, and to determine the root cause of data issues



- Facilitates definition of data standards
- Develops and maintains the repository of data standards
- Monitors and measures application compliance with data standards
- Facilitates root cause analysis of data issues
- Identifies business improvement opportunities related to fixing data issues

- Determines key attributes & business rules
- Identifies SMEs to engage
- Scopes data improvement

- Approves and prioritizes projects

Development Teams

- Provides technical application subject matter expertise to define data standards, and to determine root cause of data issues
- Ensures compliance with standards

Data Quality Scorecard

DATA SUBJECT	CSMS	EHRMS	ERP	GSDB	X.500	NOL	PSA	SPOT	TEAM WEB	WOT	TOTAL
CUSTOMER	94.26 %		98.59 %				81.51 %		99.71 %	99.94 %	**97.37 %**
TEMP CUSTOMER	7.15 %										**7.15 %**
EMPLOYEE	73.67 %	99.86 %	98.42 %	99.36 %	99.54 %		94.00 %	99.96 %	99.92 %	99.92 %	**99.20 %**
NON EMPLOYEE		93.73 %	99.48 %		69.37 %		82.92 %	99.97 %	99.95 %	99.95 %	**86.53 %**
PRODUCT	98.74 %		98.55 %	99.97 %		99.28 %		99.96 %		99.90 %	**99.83 %**
OPPORTUNITY									99.93 %	99.40 %	**99.93 %**
QUOTE										99.88 %	**99.88 %**
ORDER			98.90 %								**98.90 %**
INVOICE			99.48 %								**99.48 %**
INSTALLED BASE			71.93 %								**71.93 %**
CONTRACT	99.15 %										**99.15 %**
TOTAL	**95.22 %**	**99.43 %**	**99.20 %**	**99.97 %**	**90.90 %**	**99.28 %**	**89.46 %**	**99.96 %**	**99.89 %**	**99.89 %**	**99.30 %**

The MIT Information Quality Industry Symposium, 2007