



**ORACLE<sup>®</sup>**

## **Oracle Enterprise Data Quality for Data Integrator**

**Ljiljana Perica, Oracle Business Solution Team Leader**

**[ljiljana.perica@oracle.com](mailto:ljiljana.perica@oracle.com)**

# A World Without Data Quality

...data quality concerns are a part of your everyday world

Q: Why is my marketing campaign conversion rate so low?



## A: Data Quality

Many times large retailers maintain separate mail marketing lists which contain similar but slightly different data, without the data quality intelligence to match the duplicates.

Q: Why don't the product guys realize that garbage data in is garbage data out?



## A: Data Quality

For an IT organization, the Data Quality Service can provide an automated way to do bulk cleansing of bad data – improving data quality with minimal human intervention required

- Data are of high quality "if they are fit for their intended uses in operations, decision making and planning" (J. M. Juran)
- DQ has direct and relevant impact on processes, costs for all goods and services for the final customer/user.
- Has a direct impact on DSS for all companies.
- In some application environment such as DW, custom DQ framework can represent up to 50% of DW project costs.
- Non qualified data are basically non reliable data.

# What is Data Quality?

1. Degree of Excellence of the Data
2. Process of making and keeping data in a state of completeness, validity, consistency, timeliness and accuracy that makes it appropriate for a specific use.

# Example of Data Quality Issues

## Typical Customer Data Problems

Inconsistent formats

Abbreviations  
(often ambiguous)

Attributes non-standard,  
missing or invalid

Customer ID	Customer Name	Address 1	Address 2	City	State	Zip	Country	Birth Date	Gender
AD23298	Mr Peter Mayhew	9407 Main St		Fairfax	VA	22031-4001	USA	02/23/61	M
VS38611	Dr Ellen Van Der Heijde <b>1</b>	144 E Grove St		Kingston	PA	18704	US	07/12/57	
DC18223	Jalila Abdul-Alim (Do Not Call) <b>2</b>	4548 Pennsylvania Ave	Apt 205	Kansas City	MO	64111-3349	USA	02/23/63	F
CO9387A	Tayside Computers Inc <b>3</b>	4912 E 41st N		Idaho Falls	ID	83401	USA	31/03/2007 <b>7</b>	
TZ35019	Mr Zachary P Jahn	98-1731 Ipuala Loop	Aiea	Hawaii	96701	1710 <b>5</b>	United States	06/12/86	Male
CB27843	Mrs Edith Y Baba Junior	Baba Real Est. Corp. <b>3</b>	209 Stony Point Trl	Webster	NY		USA	11/17/1971	M <b>6</b>
OX80306	Andrew & Mary Baxter <b>4</b>	14 Oxbridge Way		Milford <b>6</b>	NH	03055-1614	US	05/28/67	F
JP70210	Mr RJ & Mrs FB MacDonald <b>4</b>	57 Hadleigh Close	Westlea	Swindon SN5 9BZ	MA	-	USA	-	F <b>6</b>
RD48107	Mr Andy Baxter	14 Oxbridge Wy		Milford	NH	3056	USA	01/01/01 <b>8</b>	M

Widespread duplication  
(often hard to spot)

**1** Compound Names

**2** Embedded Additional Information

**3** Mixed Business & Personal Names

**4** Multiple Names

**5** Mis-Fielded Data

**6** Erroneous Data

**7** International Date Formats

**8** Default or Dummy Data

A Simple Customer Table Sample

# What does Oracle Enterprise Data Quality for Data Integrator Includes?

- Profiling
  - Discover and assess data quality
    - Data Structure
    - Data Content (values, statistics, frequencies, ranges)
    - Data Relationships (dependencies, keys, joins)
- Quality
  - Integrity & Cleansing– Conform data to simple rules
  - Parsing & Standardization – Conforms data to complex rules
  - Matching – Detect, link and merge similar data
  - Enrichment – Set names and addresses to postal standards

# Built for the Business User

- Short learning curve & time-to-value
- Solution for owners of the business problem
- Integrated team collaboration

The screenshot displays the Oracle Data Director interface for profiling 'PAS Customers'. The main workspace is divided into several panes:

- Project Browser:** Shows a tree view of projects and data stores, including 'Profile PAS Customers'.
- Profile PAS Customers:** A central workspace showing a workflow of profiling tasks: Quickstats, Data Types, Frequencies, Patterns, Record Completeness, Max/Min, Lengths, and Characters. Below this is a 'Relationship Checking' section with tasks for Equal Attributes, Contained Attributes, and Product Relationship Check.
- Tool Palette - Profiling:** A sidebar containing various profiling tools like Character Profiler, Contained Attributes Profiler, Data Types Profiler, Date Profiler, Equal Attributes Profiler, Frequency Profiler, Length Profiler, and Max/Min Profiler.
- Results Browser - Quickstats:** A table showing summary statistics for various input fields.
- Quickstats - Summary statistics view (Uniqueness):** A horizontal bar chart showing the count of duplicates (red) and singletons (cyan) for each input field.

Input Field	Record Total	With Data	Without Data	Singleton
F_NAME	1000	1000	0	197
S_NAME	1000	1000	0	332
ADDRESS_1	1000	996	4	560
ADDRESS_2	1000	942	58	51
CITY	1000	999	1	19
COUNTRY	1000	999	1	11
POSTCODE	1000	1000	0	106
COUNTRY	1000	1000	0	0
TELEPHONE1	1000	379	621	348

Quickstats - Summary statistics view (Uniqueness)

Count: 0, 250, 500, 750, 1,000

Legend: Duplicates (Red), Singleton (Cyan)

# For More Information

Oracle Enterprise Data Quality



[http://www.oracle.com/us/products/middleware/  
data-integration/enterprise-data-quality/overview/index.html](http://www.oracle.com/us/products/middleware/data-integration/enterprise-data-quality/overview/index.html)

**ORACLE®**