

# De-duping Your Database: A Step-by-Step Guide

Featured Speaker: Chris Weld, Technical Solutions Consultant, Experian QAS

---



Teleconference #: 1-888-732-6202      Passcode#: 195121

# Today's Agenda

---

- Why you should have a data quality strategy
- How duplicate records affect your database
- Tutorial of QAS matchIT
  - ◆ Step-by-step run through of the duplicate identification process using QAS matchIT
  - ◆ Highlight customizable features of the tool
- Q&A

# Many Ways Data Quality Impacts An Organization

---

- Ability to market, cross-sell
- Timely delivery of products and services
- Bills collection
- Wasted printing and postage costs
- Efficiency of call center personnel, back-office staff
- Prompt customer service

Good Contact Data =  
Having the right pieces of contact information  
to enable a good relationship with your customers and prospects

# Implementing a Data Quality Strategy

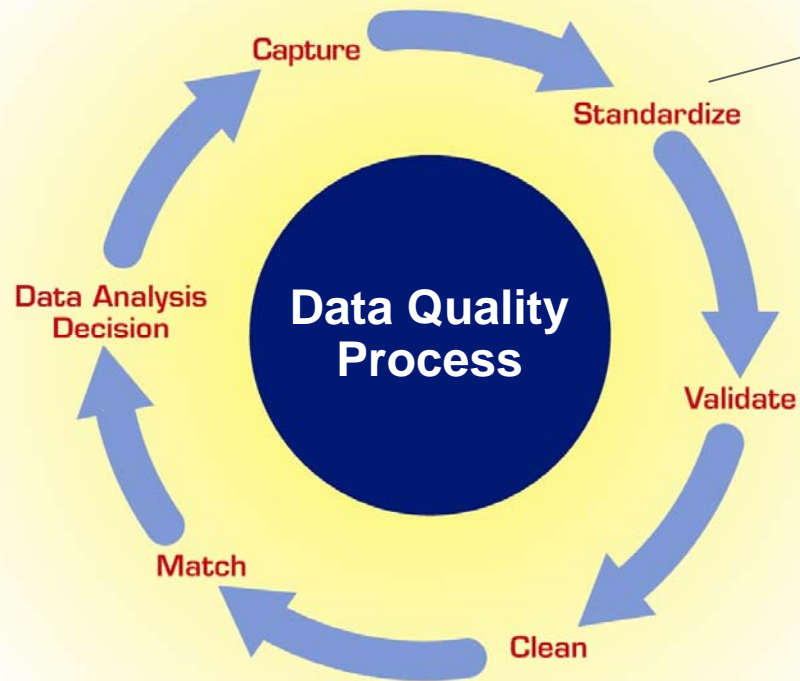


# Implementing a Data Quality Strategy

Accurate  
from  
the start

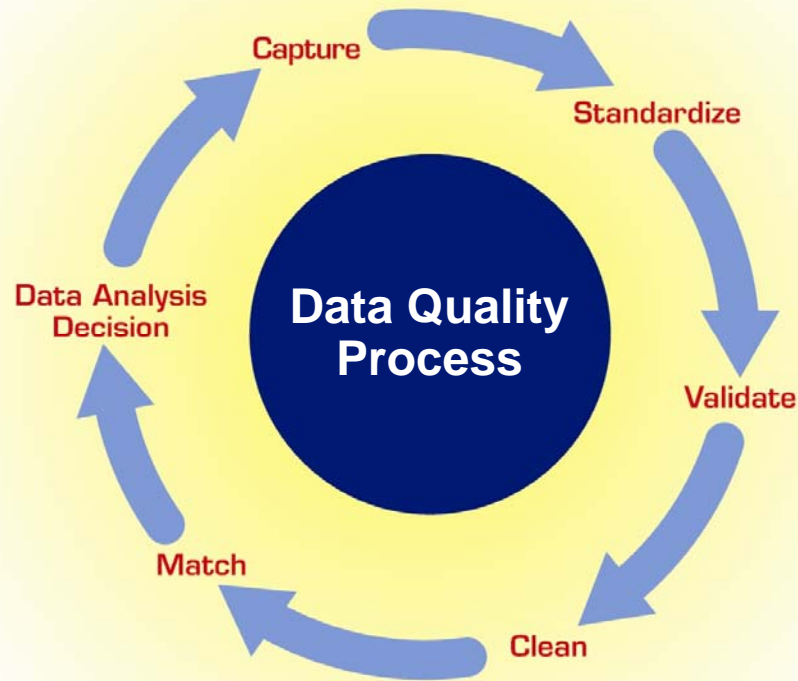


# Implementing a Data Quality Strategy



All about consistency

# Implementing a Data Quality Strategy



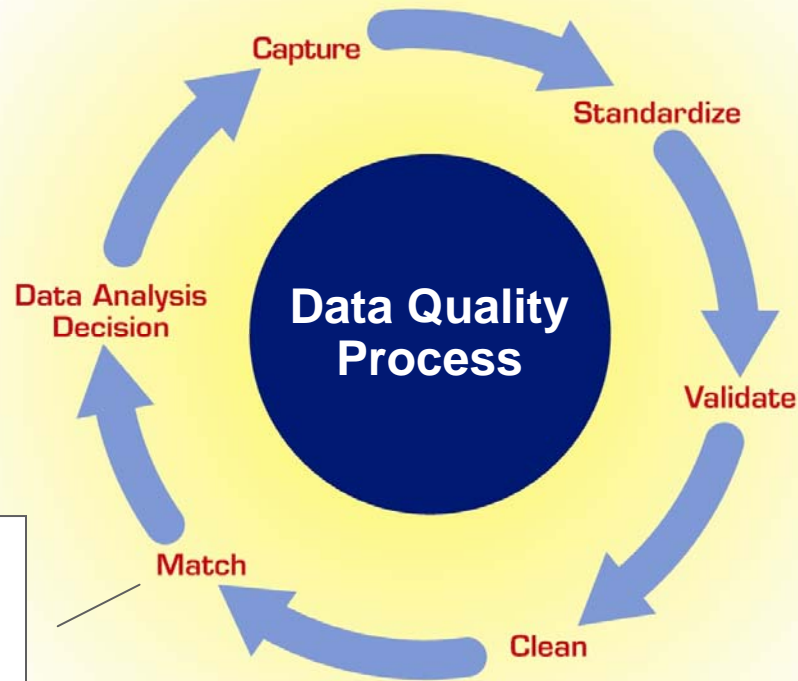
Is it complete?  
Will it get there?

# Implementing a Data Quality Strategy



Regular  
Maintenance  
Not a “one-off”

# Implementing a Data Quality Strategy



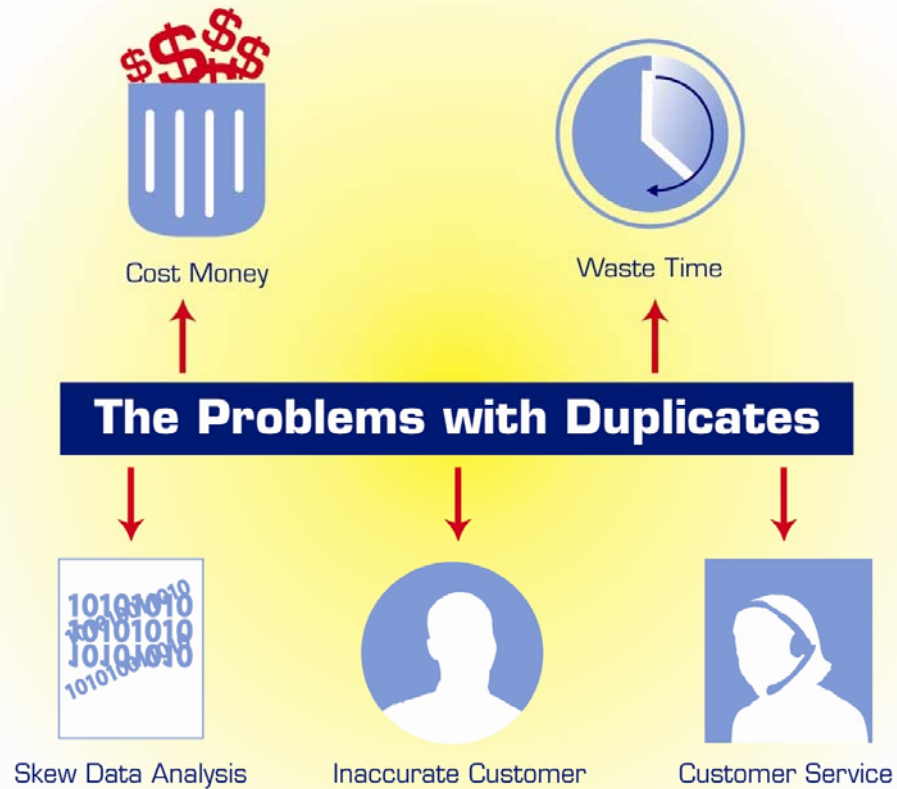
Ensure one view of the customer

# Implementing a Data Quality Strategy

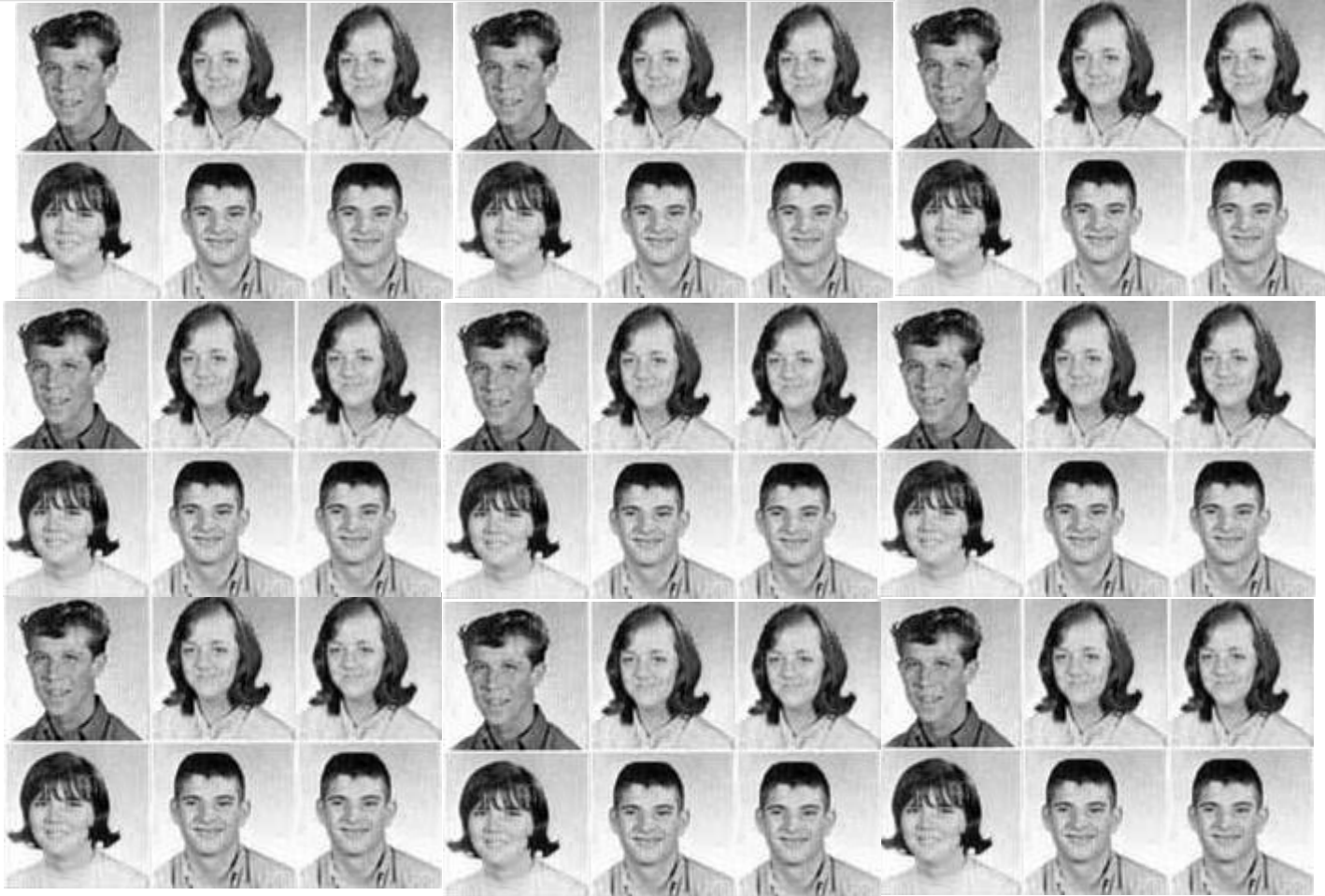
Do you trust your data?



# Today's Focus: Effects of Duplicates



# The step-by-step process to remove duplicate records from your database



# Step 1: Understand your database

---

- Take a look at your database and identify:
  - ◆ What information you take in
  - ◆ How often you take that information
  - ◆ How that information is formatted

## Step 2: Pull your data file

---

- Decide on what data you want to de-dupe
  - ◆ Which elements are you looking at
  - ◆ Entire database vs. segmentation
- Level of de-duping required for your organization
  - ◆ EX: Jennifer Smith and David Lang addresses are both 1 River Rd
- Be aware of format

## Step 3: De-duplicate file

---

- Ways to de-dupe:
  - ◆ Manual/Visual
  - ◆ DBA queries
  - ◆ Software/Services
    - ▶ Techniques
      - Phonetic: Dayton=Deighton
      - Acronym/Initial matching: “Cable News Newtwork”=“CNN”
      - Character occurrence: Wilson=Wislon
      - Table based: William=Bill=Will=Billy
      - Element matching: Mr. J. Smith=John Smith=Smith John

## Step 4: Import de-duped file

---

- Merge/Purge data
- Get the clean information back in your database
  - ◆ Import the information through the appropriate channels depending on the identification method you choose
    - ▶ Methods of import
      - ODBC
      - CSV
      - XLS

## Step 5: Continue de-dupe maintenance

---

- Importance of regularly de-duping of your database
  - ◆ Save money
  - ◆ Understand customer
  - ◆ Satisfied customers
  - ◆ Better utilize time
  - ◆ Environmentally friendly

**matchIT**

The logo graphic consists of four circles arranged horizontally. The first, second, and fourth circles are solid red. The third circle is a red outline, representing a missing or unknown value.

# Questions:

---

- Call 888-727-3985
- Email [us.webinars@qas.com](mailto:us.webinars@qas.com)
- Visit [www.qas.com/matchit](http://www.qas.com/matchit)
- New Customer Community:  
[www.qas.com/customers](http://www.qas.com/customers)





Experian™

A world of insight