

# Module 4

## Data Integration Concepts, Processes, and Techniques

### Lesson 5: Matching and Consolidation



# Lesson Objectives

- Discuss an important application for entity matching
- Explain the need for labor intensive investigations
- Reflect on the relationship between entity matching and consolidation



# Entity Matching

- Identify common entities from separate data sources
- Difficult matching process: no common identifier
- Data mining problem
  - Also known as the record linkage, entity identification, and entity resolution
  - Many approaches
  - Improve data quality for better matching results



# Matching Example

Source 1	
First name	Aimee
Middle name	Christina
Last name	Parker
Job title	Product Manager
Firm	Microsoft Corporation
Street	15580 NE 31st Street
City	Redmond
State	WA
Postal Code	98052
Country	USA

Source 2	
First name	Aimee
Middle name	C.
Last name	Parker-Lewis
Job title	Prod. Mgr.
Firm	Microsoft
Street	16517 78 <sup>th</sup> Place NE
City	Bothell
State	WA
Postal Code	98020
Country	USA



# Entity Matching Applications

Marketing



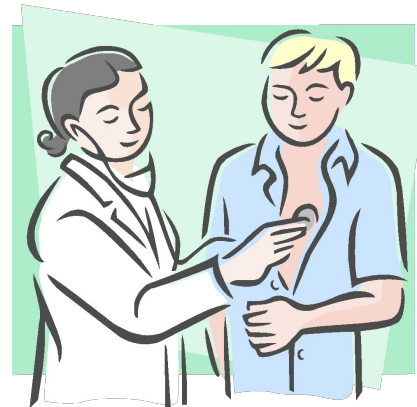
Law enforcement



Fraud detection



Health care



# Entity Matching Outcomes

<b>Predicted</b>	<b>Actual</b>	
	<i>Match</i>	<i>Non Match</i>
<i>Match</i>	True match	False match
<i>Possible Match</i>	Investigation	Investigation
<i>Non Match</i>	False non match	True non match



# Consolidation

- Merging matched records
- Linking matched records
  - Households
  - Transactions





# Merging Example

Target	
First name	Aimee
Middle name	Christina
Last name	Parker-Lewis
Job title	Product Manager
Firm	Microsoft Corporation
Street	16517 78 <sup>th</sup> Place NE
City	Bothell
State	WA
Postal Code	98020
Country	USA





# Household Consolidation



**George  
Smith**



**Janet  
Smith**



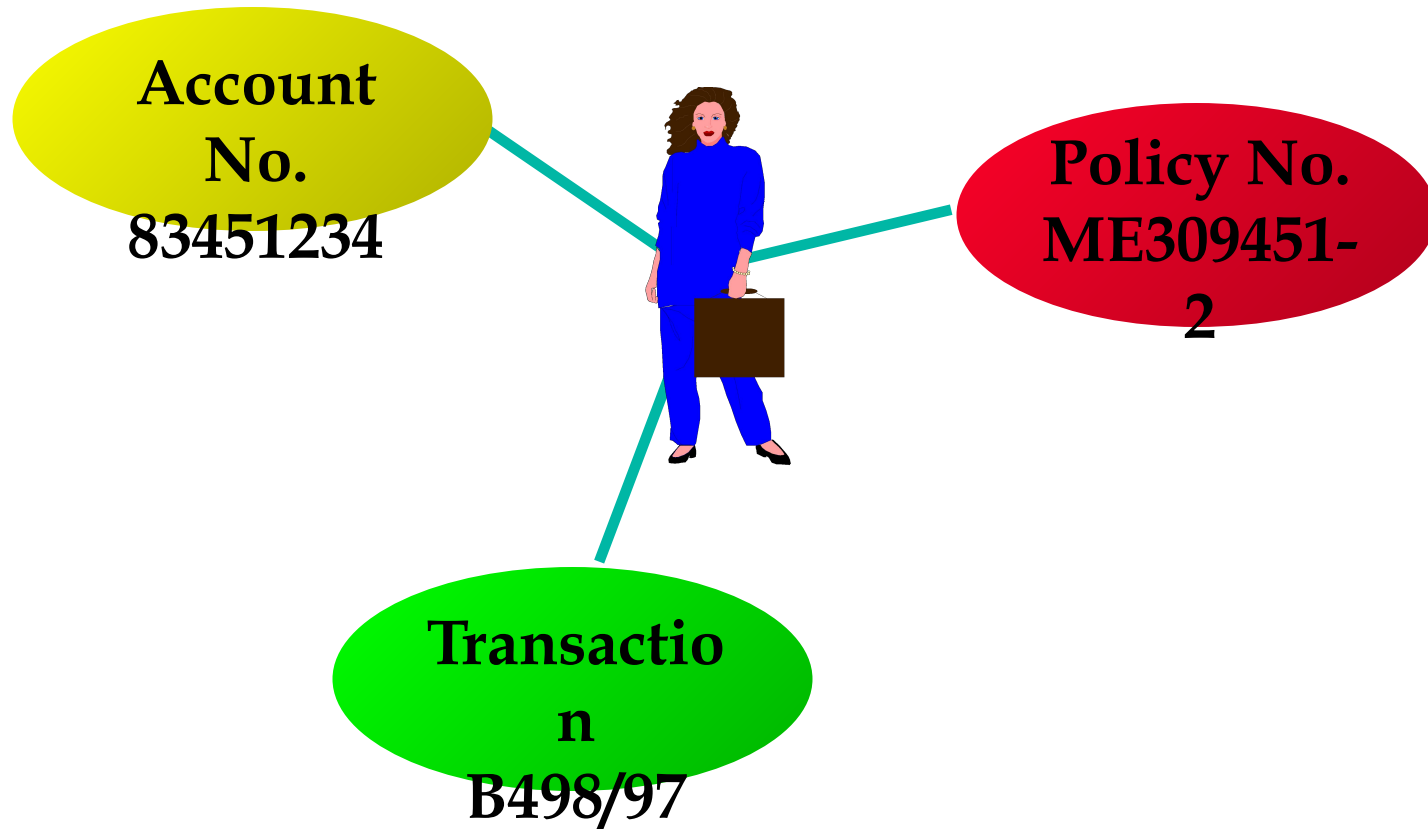
**Karen  
Smith**



**Thomas  
Smith**



# Transaction Linking



10



# Summary

- Importance and difficulty of entity matching
- Costly disposition of possible matches
- Consolidate by merging and linking

