

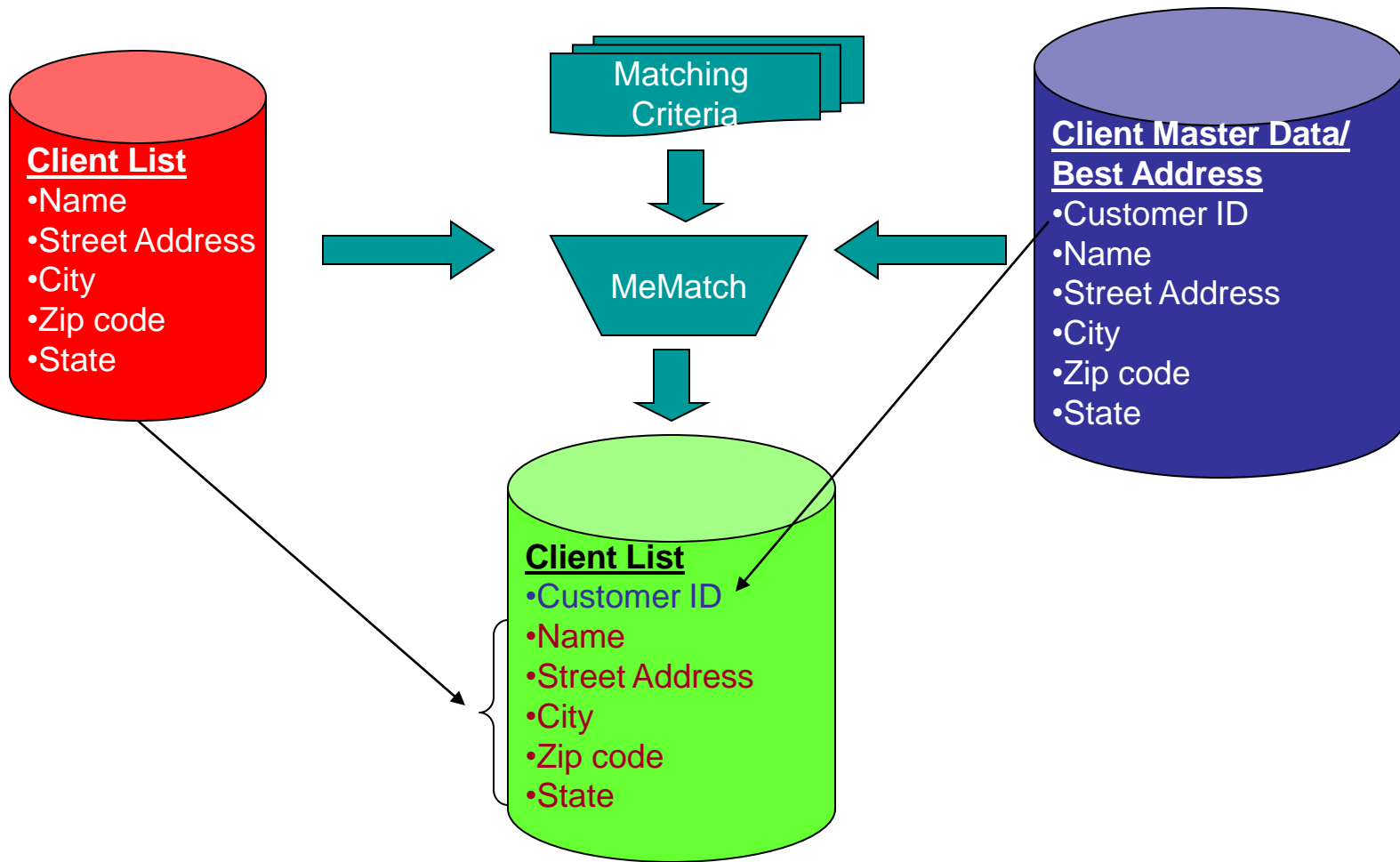
Data Means MEMatch Algorithm

Business Case

- Client has a file with customer contact info that needs to be cross referenced against another file having a field of interest such as “Customer ID#”.
- Since contact info is not kept in a standard format and may have slight variations in both files, a simple match by “Name”, “Address”, “City”, “Zip Code” and “State” may produce only a few matches.
- Manual variations of “Name”, “Address”, “City”, “Zip Code” and “State” will increase # of matches but are time consuming and error prompt.

The MeMatch Solution

- Data Means has a proprietary algorithm that standardizes the data and applies variations of the matching criteria, “Name”, “Address”, “City”, “Zip Code” and “State”, to maximized the number of matches returned.



The MeMatch Solution

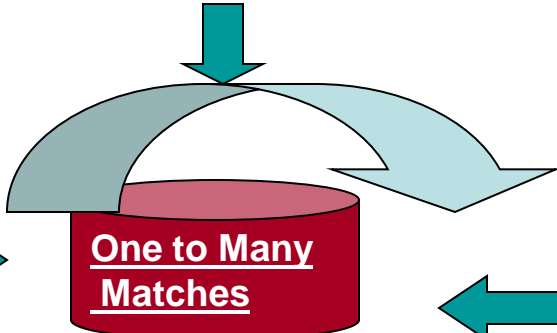
- The algorithm matches variations of contact info in both files to produce a file with the corresponding customer ID.
- The algorithm allows for the use of the sounds like criteria
- The algorithm criteria is table driven. It assumes that the criteria Rank, crit field, in the criteria table indicates the accuracy and strictness of a given criteria when compared to others. For example It assumes that crit #1 is the most accurate and strict, # 2 the second most accurate and strict.
- The algorithm goes in sequential order and loops through "n" number of matching criteria in the criteria table. A criteria table outlines the rule for matching records. For example a criteria may require exact matches on name, address, city, state and zip code. Another criteria may required matches on the first 7 characters of name, and full matches on address, city, state and zip.

Criteria Rank	Name	Addr	City	State	Zip Code
1	Full	Full	Full	Full	Full
2	10 Characters	Full	Full	Full	Full
3	Full	10 Characters	Full	Full	Full
4	10 Characters	10 Characters	Full	Full	Full
5	Full	Full	Full	Full	Full
6	10 Characters	Full	Full	Full	Full
7	Full	10 Characters	Full	Full	Full
8	10 Characters	10 Characters	Full	Full	Full
9	Excluded	Full	Full	Excluded	Full
10	Excluded	Full	Full	Full	Full
11	Excluded	10 Characters	Full	Full	Full
12	Full	Full	Excluded	Full	Full
13	10 Characters	Full	Excluded	Full	Full
14	10 Characters	10 Characters	Excluded	Full	Full
15	Excluded	10 Characters	Excluded	Full	Full
16	Full	Excluded	Full	Full	Excluded
17	Full	Full	Excluded	Full	Full
18	10 Characters	Full	Excluded	Full	Full
19	10 Characters	10 Characters	Excluded	Full	Full
20	Excluded	Full	Excluded	Full	Full
21	Excluded	10 Characters	Excluded	Full	Full
22	Full	Full	Full	Excluded	Full
23	10 Characters	Full	Full	Excluded	Full
24	Full	10 Characters	Full	Excluded	Full
25	10 Characters	10 Characters	Full	Excluded	Full
26	Full	Full	Full	Excluded	Full
27	10 Characters	Full	Full	Excluded	Full

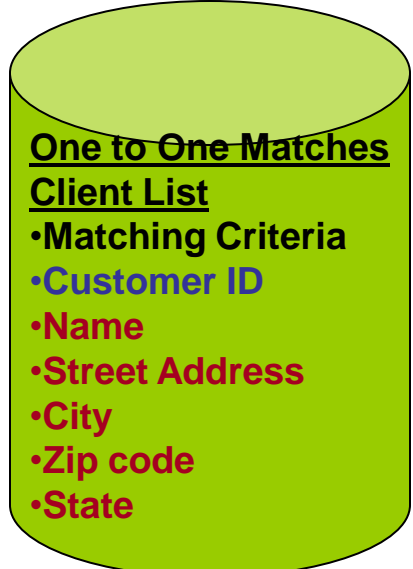
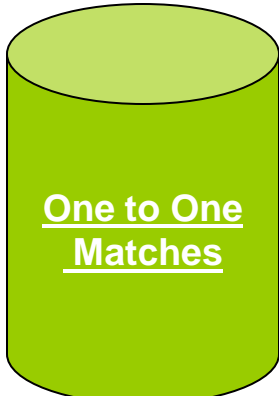
Criteria Table

- Full refers to full match on the field
- 10 characters refers to match on the first 10 characters. algorithm allows to specify # of characters to be used.
- Excluded refers to the field not being used in the matched.
- Sounds like criteria can be used.

The MeMatch Solution



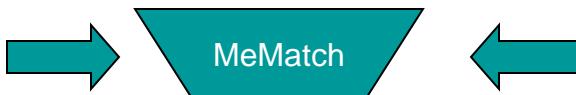
Gets Appended to "One to One" matches bucket



Resulting one to many matches and unmatches after looping through all the matching criteria



Complete Data Validation



Automated Process



Manual Validation Process

1.	One to one match based on matching criteria
2.	Unmatches
3.	One to Many Matches

- Google
- NPPEs
- Other Databases