



## Database entity matching

### The background:

A research company approached CoGo Data to match entries from two data sources. The data consisted of numbers, names, locations, and project descriptions. Data source one was data collected by the client's field workers who were tasked with taking photos of billboards. This data was populated in a structured relational database that included the numbers, names, locations, and project descriptions. Data source two was the same data contained in the client's database. The data collected by the field workers (data source one) had to be compared to the client's data base (source two) to update their database.

### The problem:

The problem was to match the entries from source one with source two. This is challenging because the information from both sources can be joined on multiple variables and these relationships are commonly many-to-many. Furthermore, these data sources are relatively large, which complicates traditional full combinatorial matching methods.

### The solution:

CoGo Data developed a solution that finds partial matches by means of natural language processing that provides the company with a matching score on all relevant variables for different matching policies (prioritisation of variables during matching).

### The result:

This automation allows the client to narrow down the search that would traditionally take weeks to perform manually to a fraction of the time, which contributes to a significant improvement in productivity.

