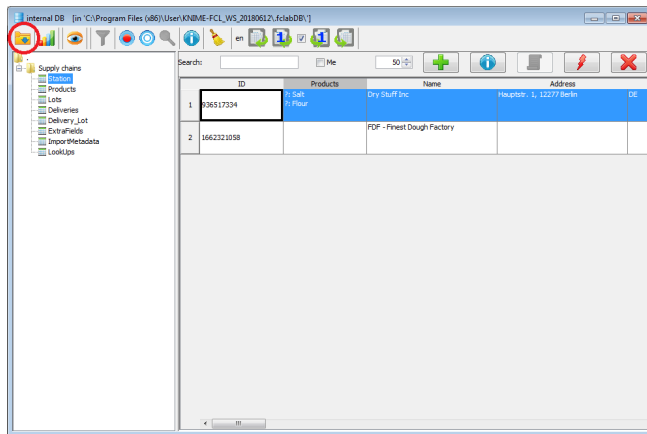# Finding and removing duplicates
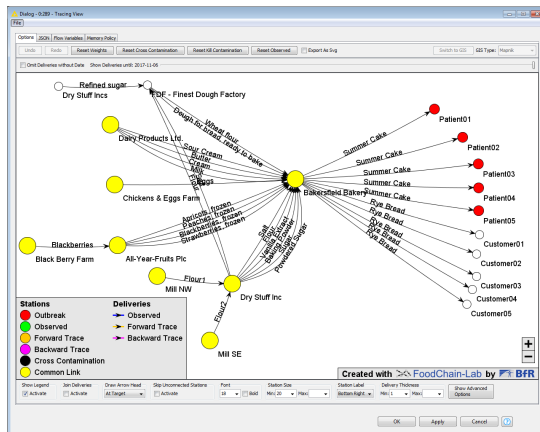
Learn

- how to find duplicates in the FoodChain-Lab database
- how to remove these duplicates
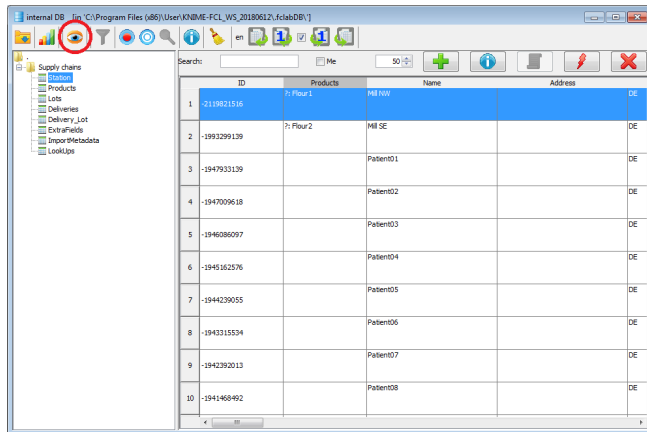
Note: You need data from the tutorials "Data collection and import" and "Tracing backward template". Please do these tutorials first.

- Open the database and import
  "FCL_Backtrace_Start_tob_en_FDF_DB_clean.xlsx".

- Have a look at the network in the **Tracing View**. There is station *Dry Stuff Inc* and another one called *Dry Stuff Incs*. In the network it looks as if the the refined sugar was delivered to the *Finest Dough Factory* by a completely different supplier. Because in reality both stations are identical this typo should be corrected.

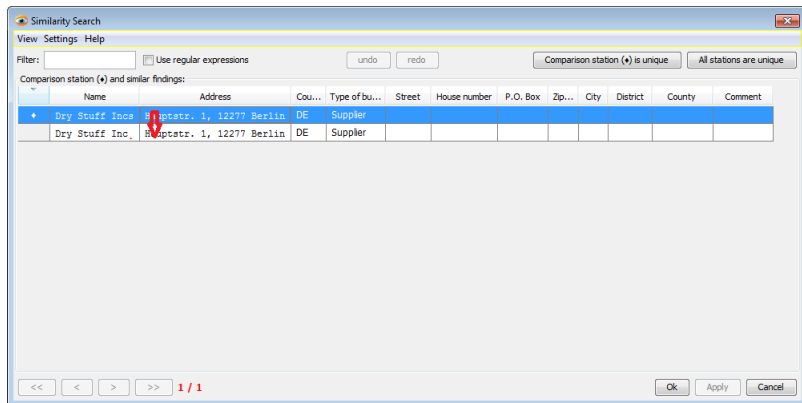- Start the similarity search (see button in the red circle).

- To search for duplicates in station names and addresses tick the encircled radio button.
- "Name: 90" means that the similarity between two or more station names must be 90 % or greater in order to be displayed as duplicate.
- Click "OK"

- Left click the row with the typo to highlight it. Then drag and drop it onto the row that should be kept. It is like moving a file (or a selection of files) to a folder in a file browser.
- Click "OK".
- Because there are no more station names with 90 % similarity the window closes.
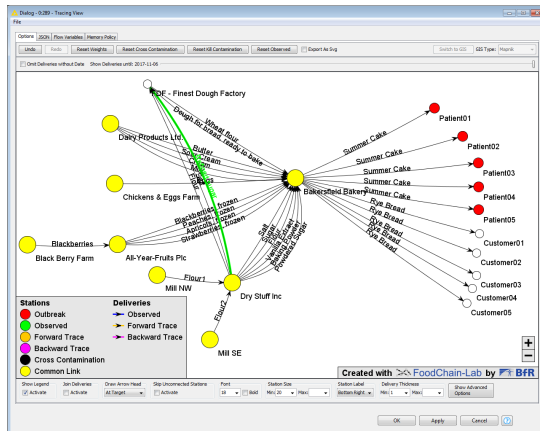
- Reset the **Supply Chain Reader** and have a look at the **Tracing View**.
- The entries have been merged. The refined sugar has been assigned to the correctly spelled company station *Dry Stuff Inc*.