

# 17 Finding and removing duplicates

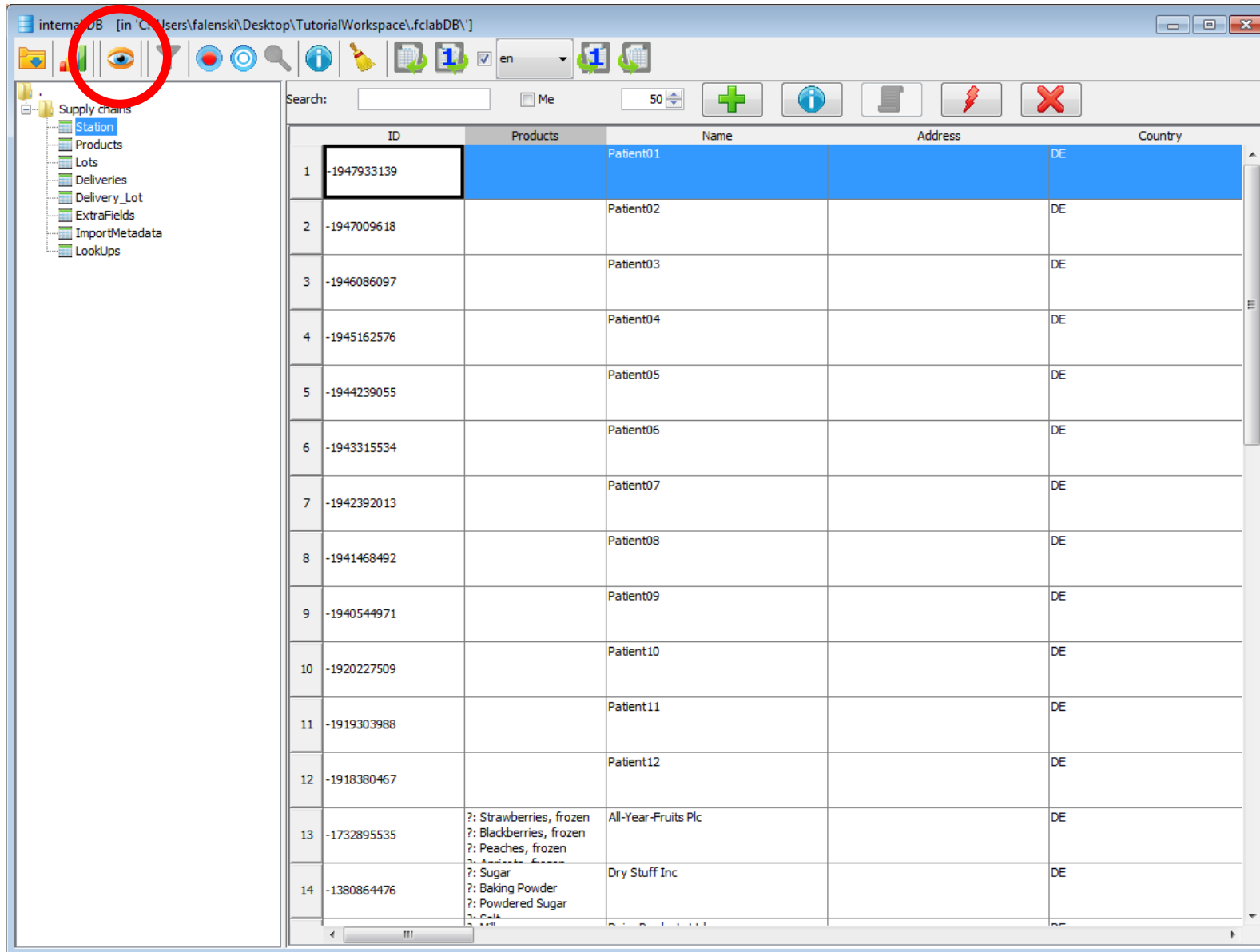


## Learn

- what is necessary to free your database from duplicates

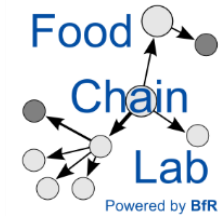
Note: You need data from the tutorials [“15 Data collection and import”](#) and [“16 Tracing Backward Template”](#). Please do these tutorials, first.

# Step 1



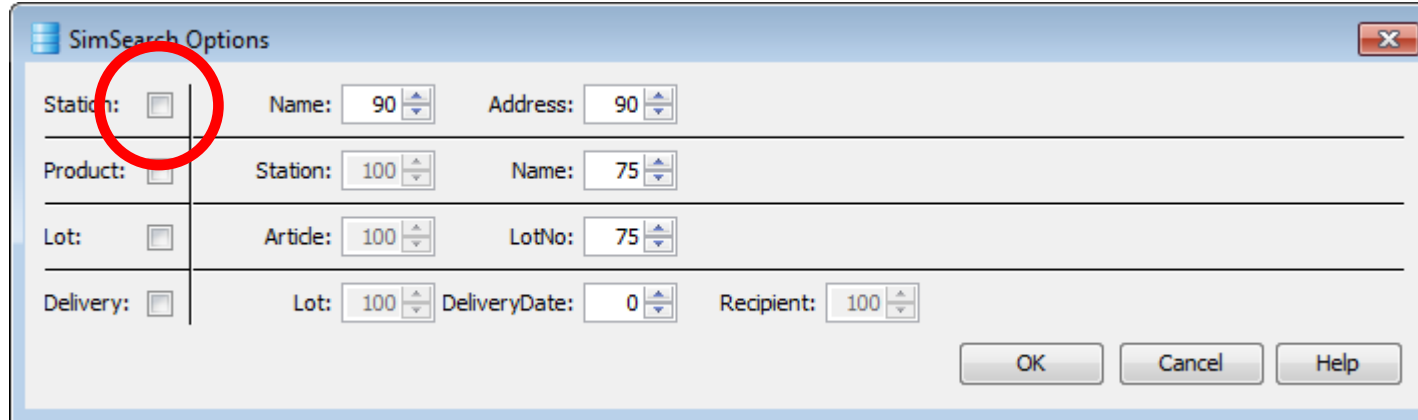
The screenshot shows a database application window titled 'internal\_DB [in 'C:\Users\falenski\Desktop\TutorialWorkspace\.fclabDB\']'. The interface includes a toolbar with various icons, a search bar, and a table of data. A red circle highlights a button in the toolbar, likely for starting a similarity search. The table has columns for ID, Products, Name, Address, and Country. The data is as follows:

ID	Products	Name	Address	Country
1	-1947933139	Patient01		DE
2	-1947009618	Patient02		DE
3	-1946086097	Patient03		DE
4	-1945162576	Patient04		DE
5	-1944239055	Patient05		DE
6	-1943315534	Patient06		DE
7	-1942392013	Patient07		DE
8	-1941468492	Patient08		DE
9	-1940544971	Patient09		DE
10	-1920227509	Patient10		DE
11	-1919303988	Patient11		DE
12	-1918380467	Patient12		DE
13	?: Strawberries, frozen ?: Blackberries, frozen ?: Peaches, frozen ?: Apples, frozen	All-Year-Fruits Plc		DE
14	?: Sugar ?: Baking Powder ?: Powdered Sugar	Dry Stuff Inc		DE



- Open the database and import „FCL\_Backtrace\_Start\_tob\_en\_FDF\_DB\_clean.xlsx“.
- Start the similarity search (see button in the red circle).

## Step 2



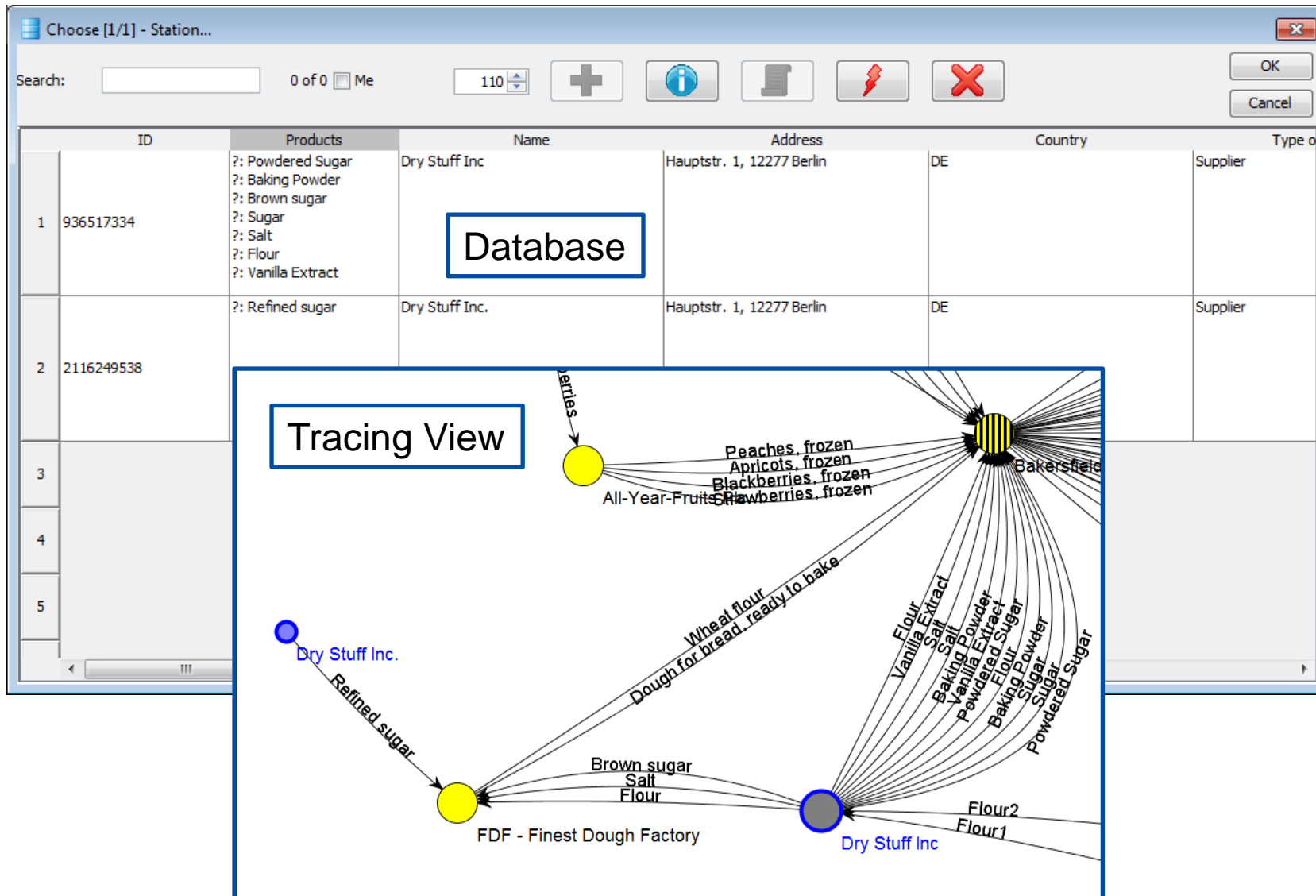
SimSearch Options

Station:	<input type="checkbox"/>	Name:	90	Address:	90		
Product:	<input type="checkbox"/>	Station:	100	Name:	75		
Lot:	<input type="checkbox"/>	Article:	100	LotNo:	75		
Delivery:	<input type="checkbox"/>	Lot:	100	DeliveryDate:	0	Recipient:	100

OK Cancel Help

- To search for duplicates in station names and addresses tick the encircled box.
- “90” means that the similarity must be 90 % or greater. If the similarity is smaller, the search will not display entries as similar.
- Click OK

# Step 3



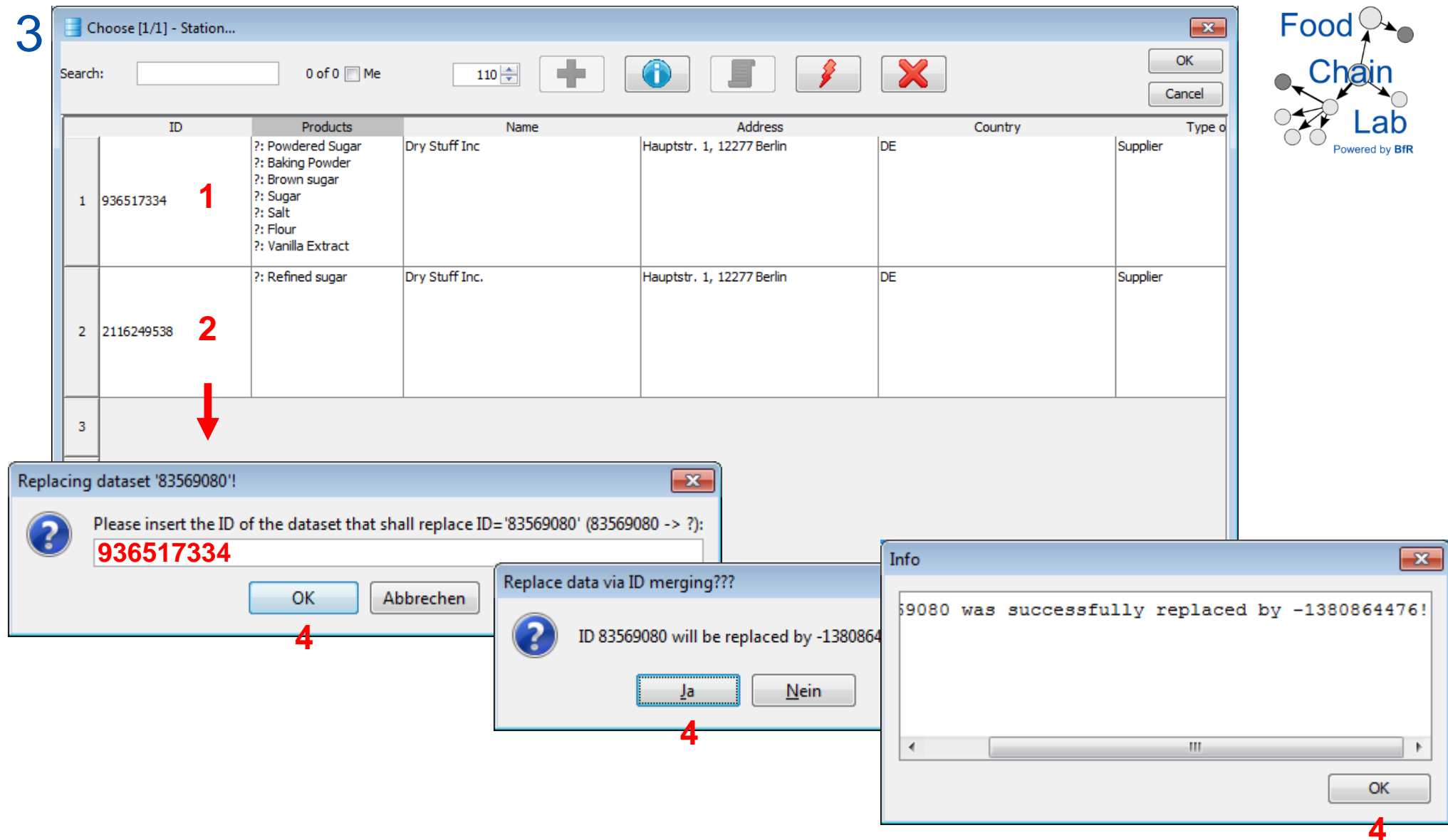
The screenshot shows a software window titled "Choose [1/1] - Station...". It contains a table with two rows of data. A "Database" label is placed over the table. Below the table, a "Tracing View" network diagram is shown, illustrating the relationships between various ingredients and suppliers.

ID	Products	Name	Address	Country	Type o
1 936517334	? : Powdered Sugar ?: Baking Powder ?: Brown sugar ?: Sugar ?: Salt ?: Flour ?: Vanilla Extract	Dry Stuff Inc	Hauptstr. 1, 12277 Berlin	DE	Supplier
2 2116249538	? : Refined sugar	Dry Stuff Inc.	Hauptstr. 1, 12277 Berlin	DE	Supplier

The Tracing View diagram shows a network of nodes and edges. A central node is labeled "Bakersfield". Other nodes include "Dry Stuff Inc.", "FDF - Finest Dough Factory", and "All-Year-Fruits". Edges represent ingredients and products, such as "Refined sugar", "Wheat flour", "Dough for bread ready to bake", "Vanilla Extract", "Salt", "Baking Powder", "Powdered Sugar", "Sugar", "Flour", "Flour2", "Flour1", "Peaches, frozen", "Apricots, frozen", "Blackberries, frozen", and "Raspberries, frozen".

There are two data rows in the database with (nearly) identical company names and with identical addresses. The network in the Tracing View does not reflect reality anymore. However, deleting one database row is not an option because different ingredients are attached to these two identical stations. We would like to merge both entries so that all ingredients come from one representation of the company Dry Stuff Inc.

## Step 3



ID	Products	Name	Address	Country	Type o
1 936517334	?: Powdered Sugar ?: Baking Powder ?: Brown sugar ?: Sugar ?: Salt ?: Flour ?: Vanilla Extract	Dry Stuff Inc	Hauptstr. 1, 12277 Berlin	DE	Supplier
2 2116249538	?: Refined sugar	Dry Stuff Inc.	Hauptstr. 1, 12277 Berlin	DE	Supplier
3					

Replacing dataset '83569080!'

Please insert the ID of the dataset that shall replace ID='83569080' (83569080 -> ?):  
**936517334**

Replace data via ID merging???

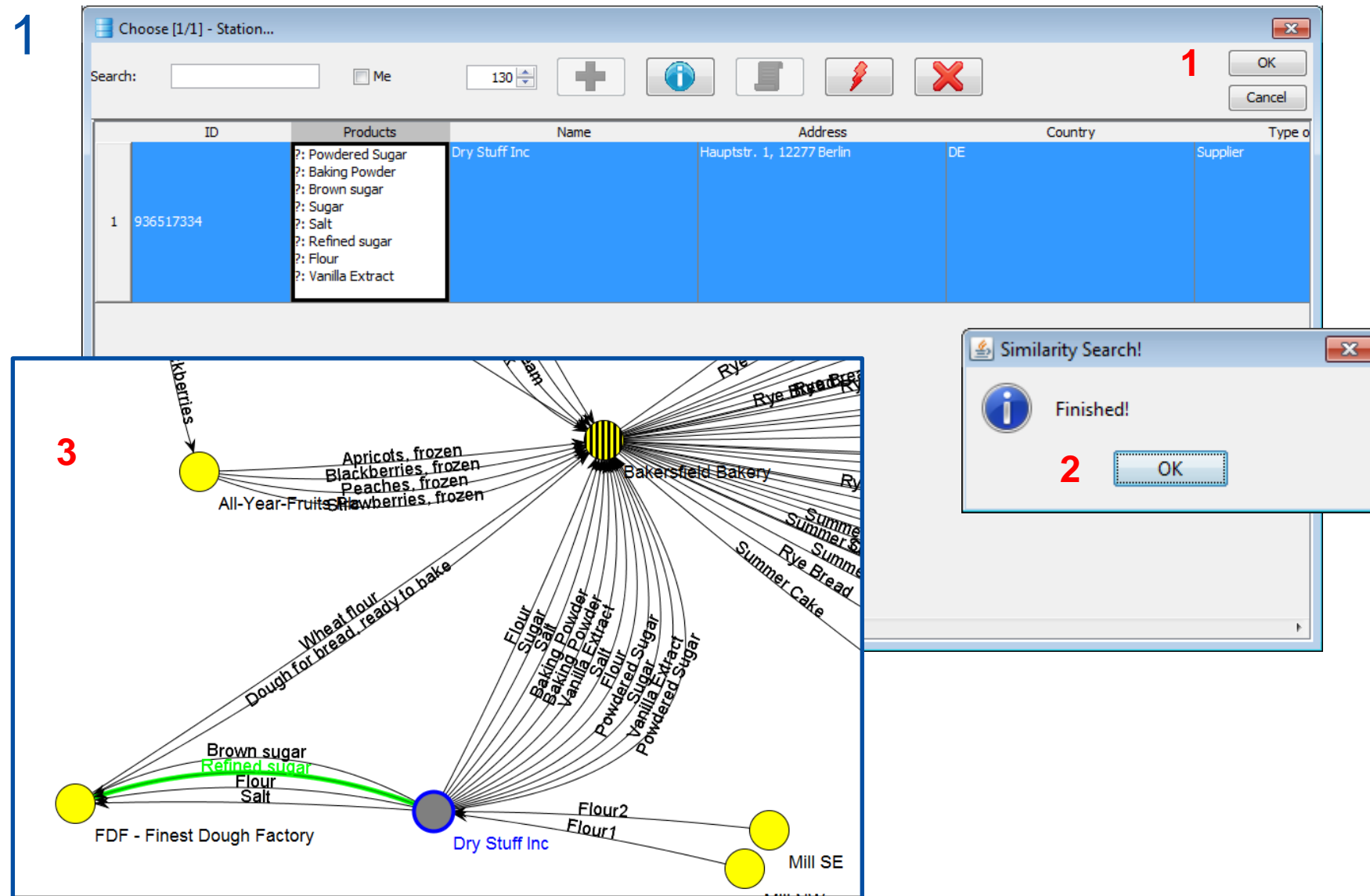
ID 83569080 will be replaced by -1380864

Info

59080 was successfully replaced by -1380864476!

- 1) Copy the ID from the entry you would like to keep (press Ctrl+C with the ID cell highlighted).
- 2) Right-click into the ID cell of the row you would like to replace.
- 3) Paste the copied ID here (Ctrl+V).
- 4) Click OK -> Ja -> OK

# Step 1



The screenshot shows a software interface with a table of products and a network diagram. The table has columns for ID, Products, Name, Address, Country, and Type. A dropdown menu is open over the 'Products' column, listing items like Powdered Sugar, Baking Powder, Brown sugar, Sugar, Salt, Refined sugar, Flour, and Vanilla Extract. A red '1' is next to the OK button. Below the table is a network diagram with nodes and edges. A red '3' is next to a yellow node. A 'Similarity Search!' dialog box with an information icon and the text 'Finished!' and a red '2' next to the OK button is overlaid on the right.

ID	Products	Name	Address	Country	Type
1 936517334	? Powdered Sugar ? Baking Powder ? Brown sugar ? Sugar ? Salt ? Refined sugar ? Flour ? Vanilla Extract	Dry Stuff Inc	Hauptstr. 1, 12277 Berlin	DE	Supplier

The entries have been merged. The refined sugar has been added to the products of the remaining database entry for the company Dry Stuff Inc.

- 1) Click OK
- 2) Click OK
- 3) Reset the Supply Chain Reader and have a look at the Tracing View.