**Loading GDPR Test Data Sets**

This document assumes that you have already installed Senzing and downloaded either the US or UK test data sets for GDPR.

You can download a free trial of Senzing from our homepage: [www.senzing.com](http://www.senzing.com)

Contents

[Step 1: Extract the Files 2](#_Toc517270757)

[Step 2: Add the Files to Senzing 3](#_Toc517270758)

[Step 3: Mapping the Employee Data Source 4](#_Toc517270759)

[Step 4: Load all the Files 7](#_Toc517270760)

[Step 5: Review the Results 8](#_Toc517270761)

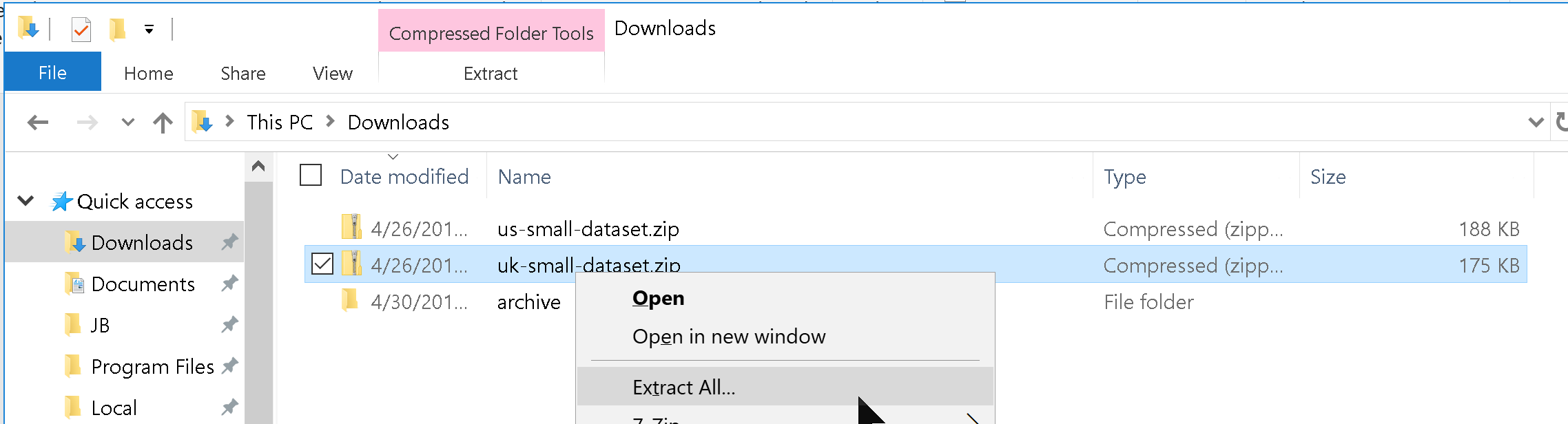
[Step 6: Review the Dashboard 10](#_Toc517270762)

[Step 7: Perform Single Subject Searches 11](#_Toc517270763)

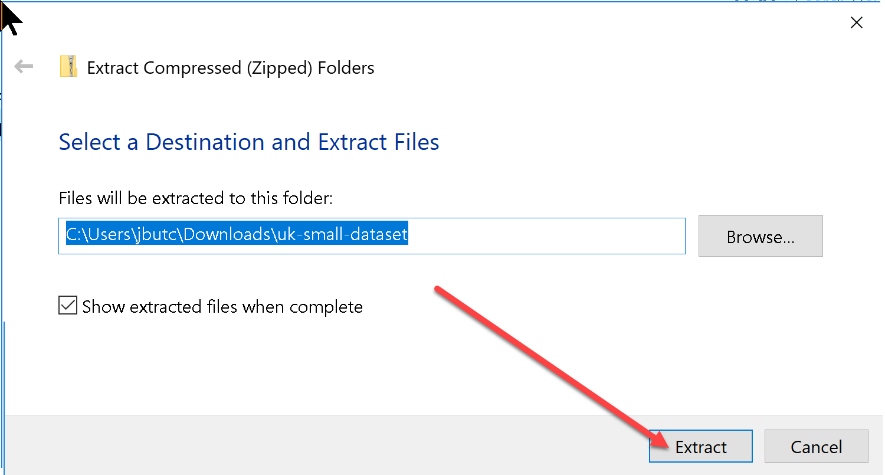
[Next Steps 13](#_Toc517270764)

# Step 1: Extract the Files

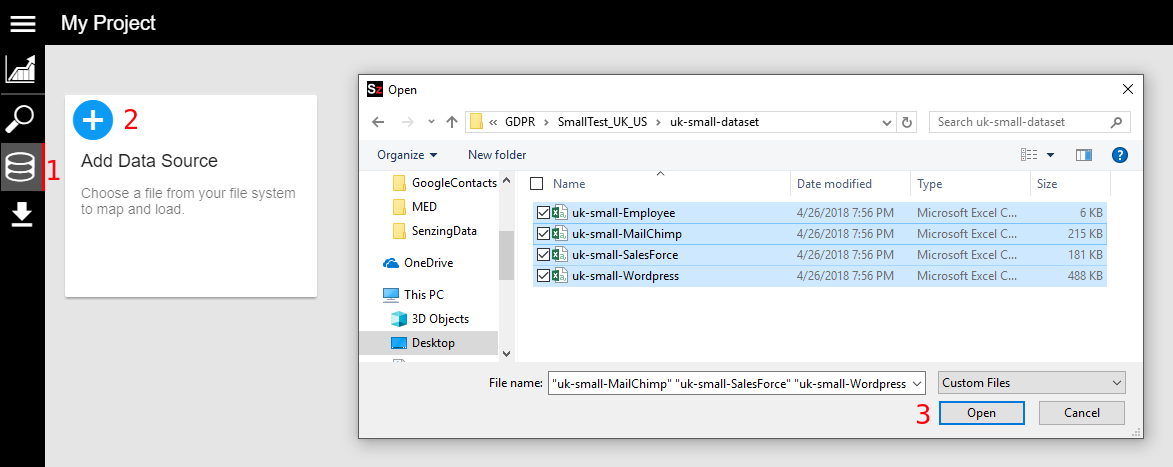
In Windows explorer, right click on the zip file and choose ‘Extract All…’



Choose the location to extract the files to.

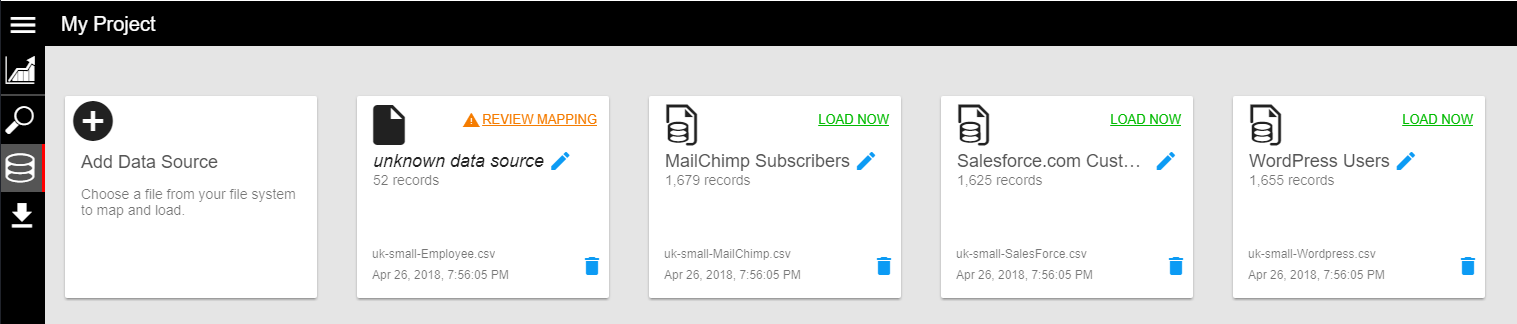


# Step 2: Add the Files to Senzing



1. On the left navigation bar click the add data icon.
2. In the file selection box, navigate to the directory you extracted the files to and select them all.
3. Press the ‘Open’ button.

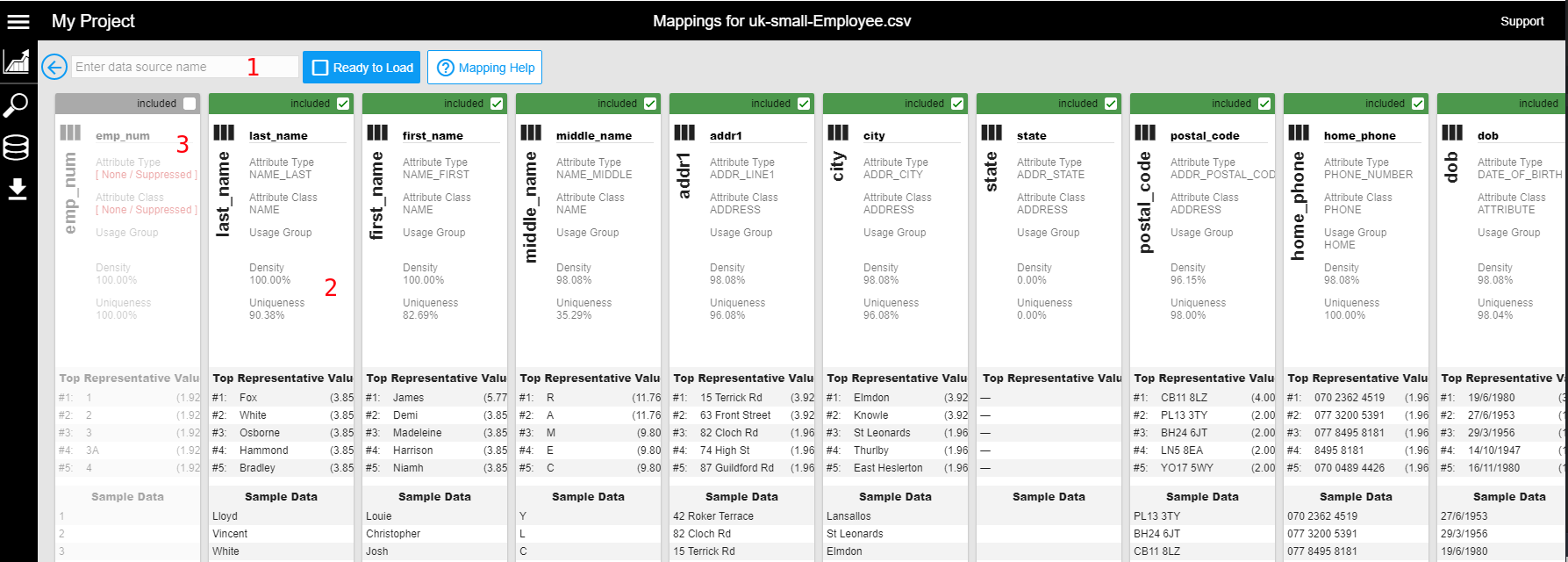
After a brief processing step, the screen will look like:

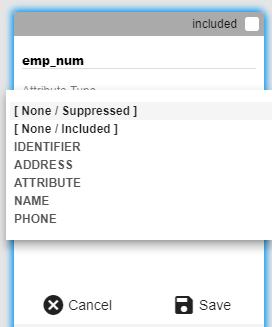
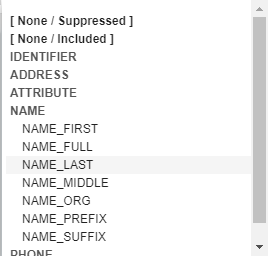


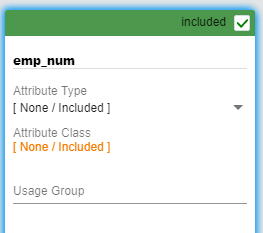
There are four files in each test data set. Three are recognized formats from widely used applications - Salesforce, MailChimp and WordPress. We purposely sent an unrecognized format for employees, to allow you to see how easy it is to map a new data source to the terms Senzing understands and uses for entity resolution.

# Step 3: Mapping the Employee Data Source

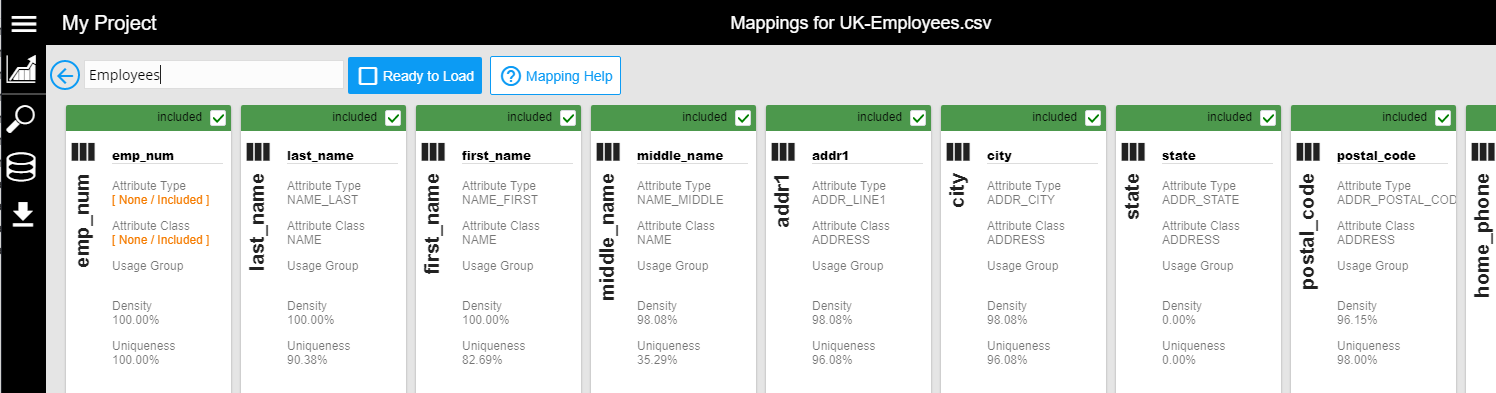
Clicking on **REVIEW MAPPING** on the employee data source card brings up the mapping screen.



1. Give your data source a name, since these are employees name the data source **Employees**
2. Review the statistics and data in each column, noticing
   * The **emp\_num** field is 100% populated and 100% unique.
   * The **last\_name** field is 100% populated but only 90% unique with Fox being the most common value.
   * What is the most common **first\_name** value?
3. Your job in mapping is assigning an attribute type to each column. Senzing automatically detects and assigns as many as it can, notice
   * The **last\_name** column was automatically mapped to the **NAME\_LAST** attribute. Do those look like last names to you?
   * The **emp\_num** column was not mapped to a recognized attribute, but it is useful data to load and make available for reference; maybe there is a way to keep it!
   * Clicking on **Attribute Type** in any column will display a list of recognized attributes.
   * Click on a section to see the attributes in it, they are self-explanatory.  
       
      
4. To load the **emp\_num** column with the mapped data without mapping it click the **Included** checkbox at the top of the column. Notice the attribute type changes to **None / Included** indicating the column isn’t mapped to be used during Entity Resolution processing but it will be loaded and available in reports and exports.
   * In addition to **emp\_num** mark **job\_title** and **emp\_status** as Included also.

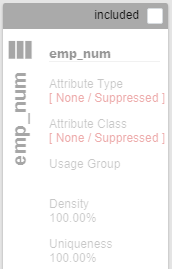


Click **Ready to Load** when you are satisfied with your mappings.

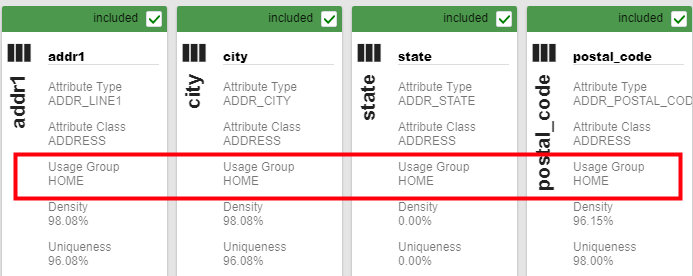


The changes that were made to complete the mapping:

1. No changes to automatic attribute assignments, they all looked good.
2. Mark **emp\_num**, **job\_title** and **emp\_status** as **Included**. Now, when presented with a match between an employee and another data source such as a SalesForce customer, you’ll observe if they are a current or past employee, their job title and unique employee number.
3. If there were any sensitive fields in the file exported from your human resources system - such as salary or ethnicity – you could choose to suppress them using the **None / Suppressed** mapping option. Shown here with **emp\_num** as an example.

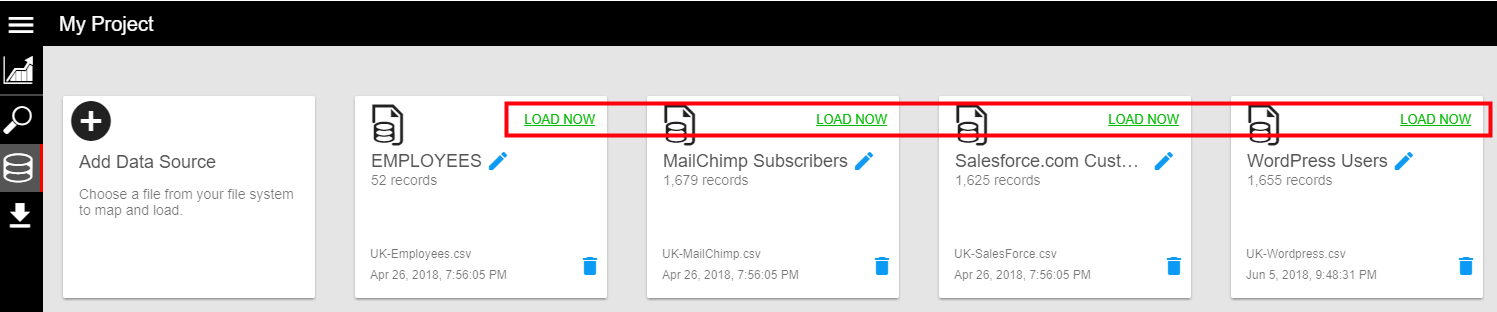


The Usage Group only needs to be set when a data source has duplicate attributes. For instance, if your source has four fields for a home address - address line 1, city, state and postal code - and another set of four for a mailing address, you would set the usage group to **HOME** or **MAILING** accordingly. This groups the fields together to inform Senzing which attributes go together to form a complete address.

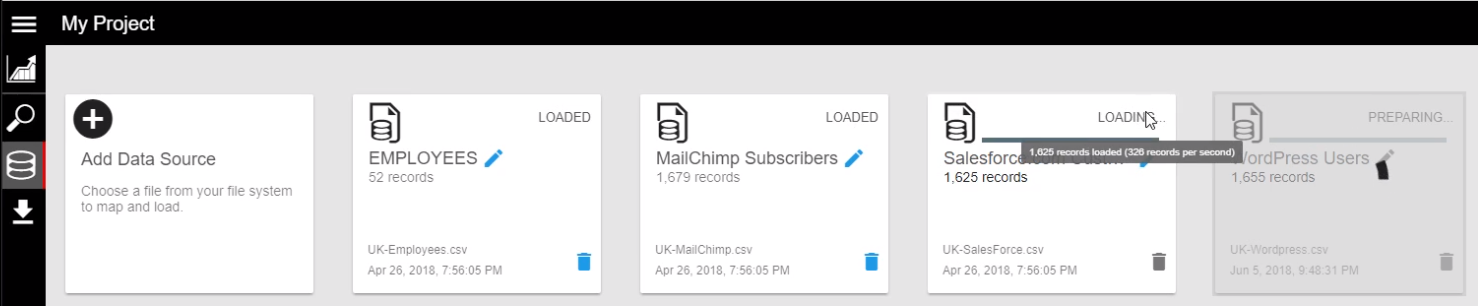


# Step 4: Load all the Files

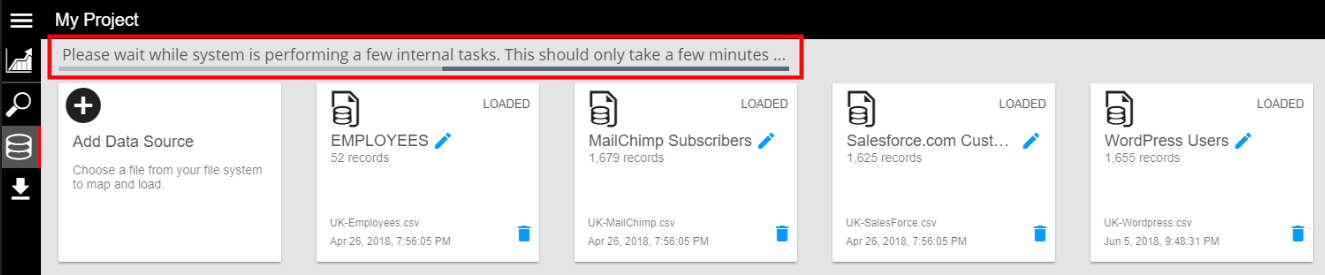
Click the load now button on all 4 without waiting for each to complete, each will be queued ready to load.



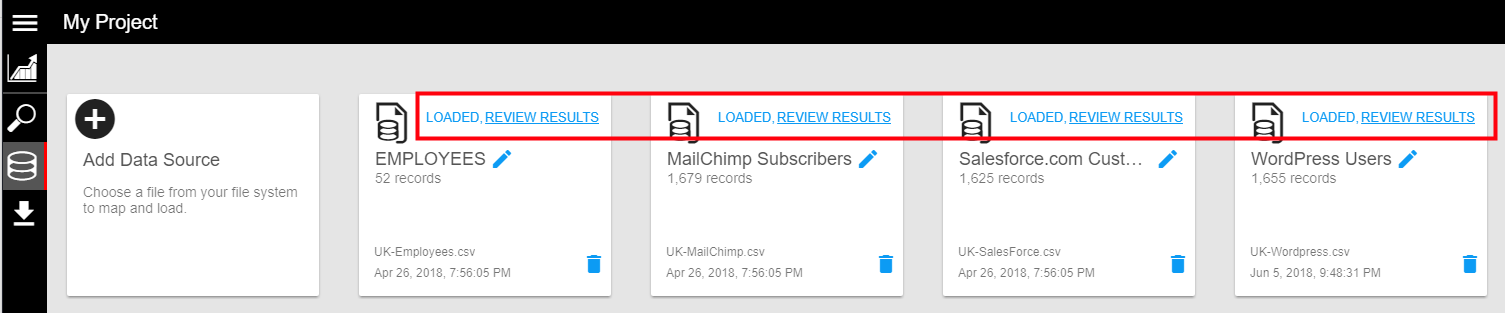
While loading, you can hover over the loading status to monitor the progress.



Occasionally you may see a message indicating the system is busy. It usually completes in a minute or two.

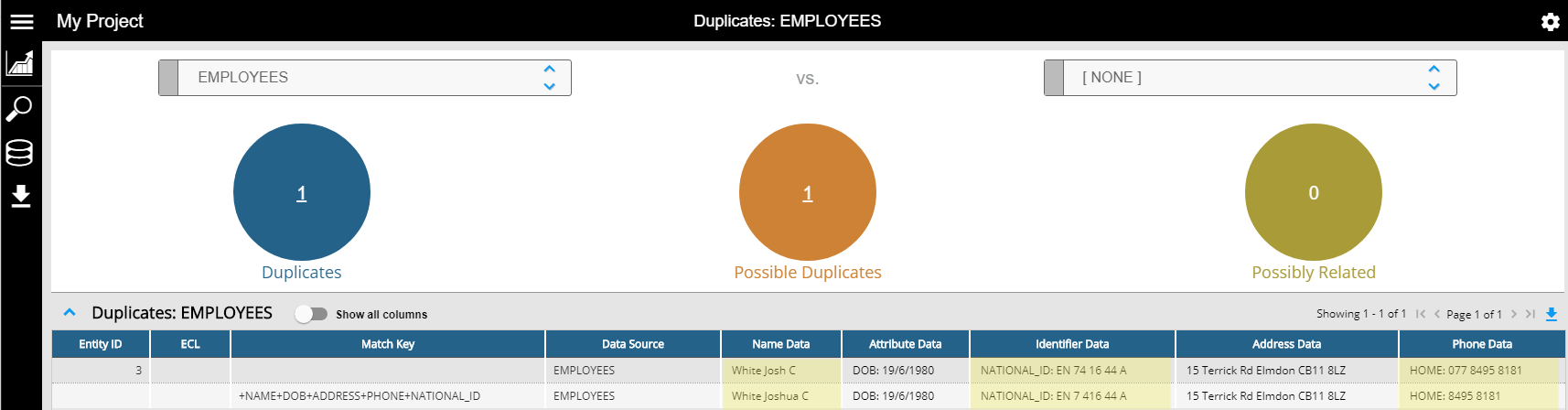


Once loaded click on **LOADED, REVIEW RESULTS** of any card to see the statistics for that data set.



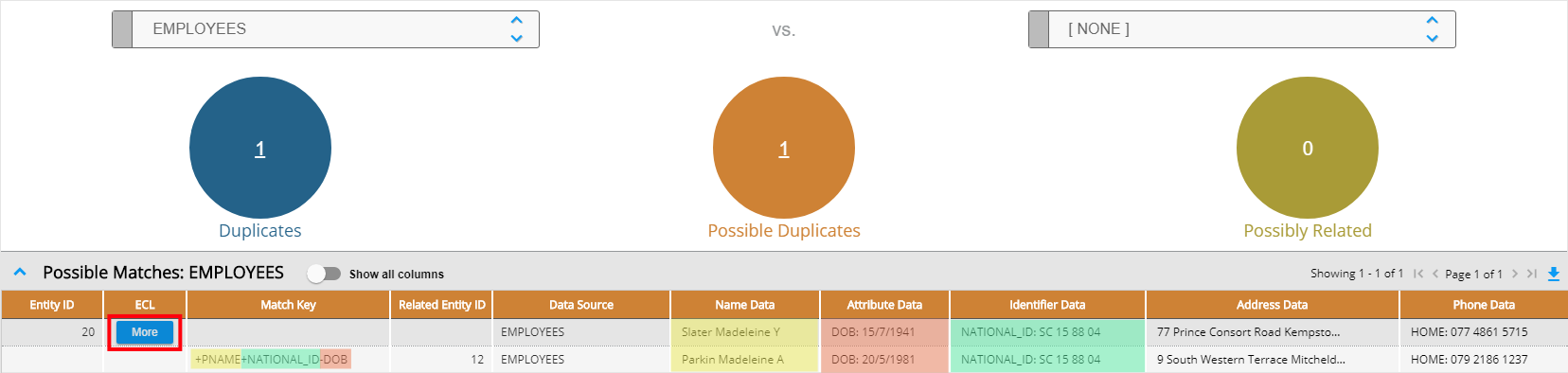
# Step 5: Review the Results

Clicking on **LOADED, REVIEW RESULTS** for the Employee data source card will take you to a review screen.

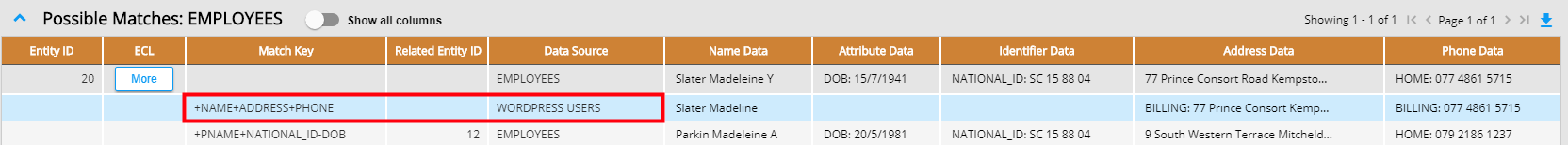


* Highlights show how we match through name variants, badly formatted National ID numbers and a partial phone number.
* There are enough matching values here to consider these two records a duplicate.

Clicking the number in the middle of the **Possible Duplicates** circle shows records determined to be possible duplicates for review.

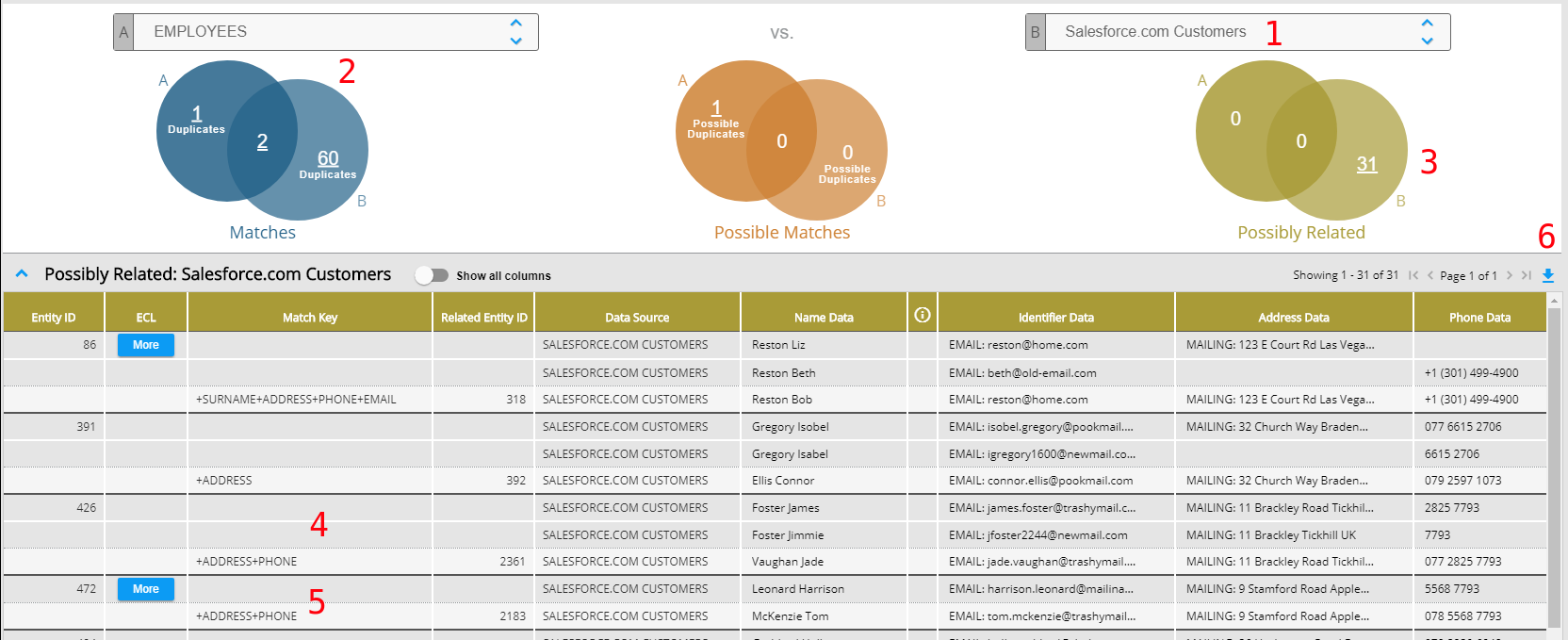


* The **Match Key** column is the ‘key’ to show what attributes added (+) and what attributes subtracted (-) from the match.
* This was downgraded to a possible match due to conflicting date of birth and only a PNAME (partial name) match.
* The **More** button in the ECL (Entity Centric Learning) column indicates there are additional records from other data sources that are part of the match and entity(s). Press the **More** button to display them.



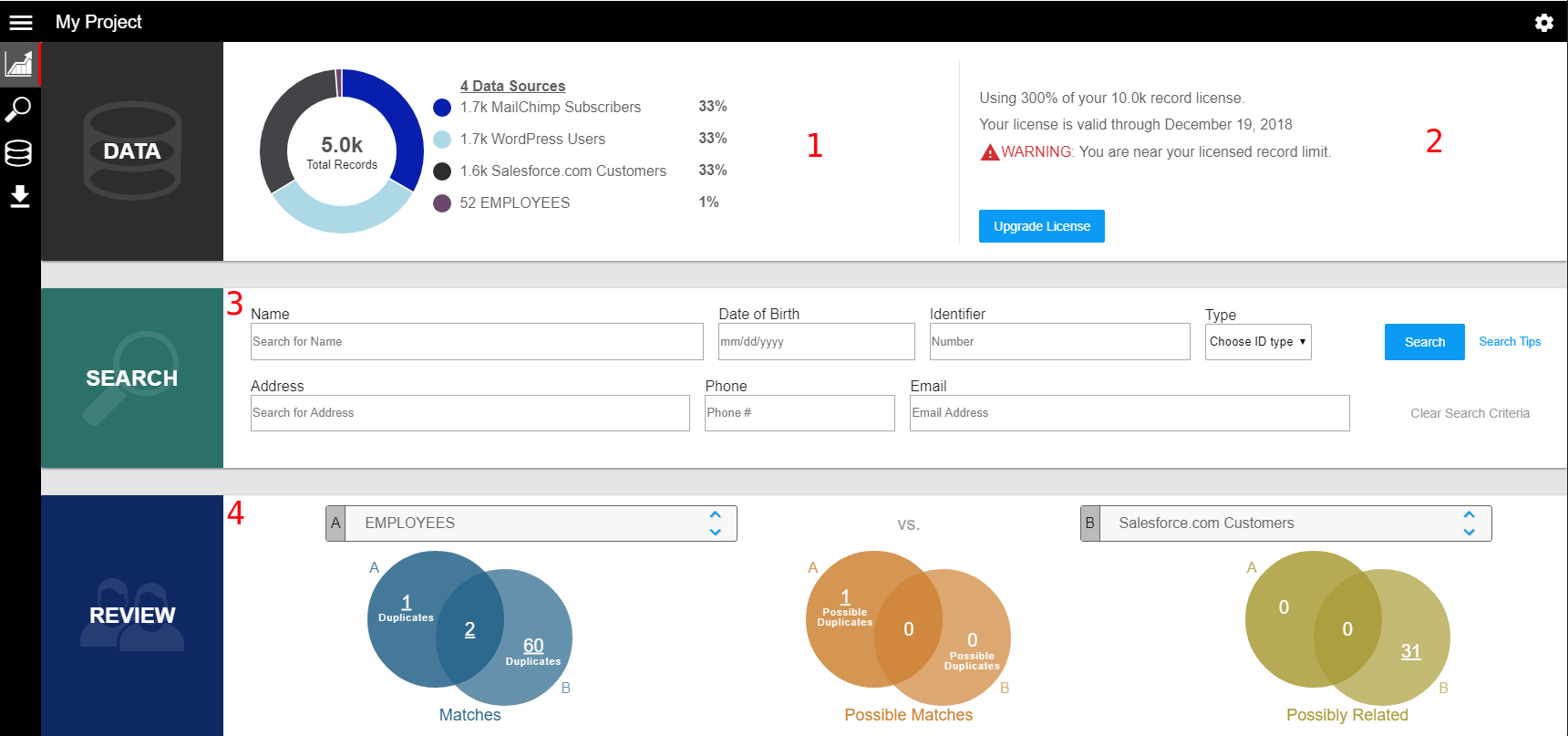
Note the new row from the WordPress Users data source. The entity identified as **Entity ID** 20 consists of 2 records – 1 from the employee data source and the other from WordPress Users. This entity has a relationship (possible duplicate) to the entity identified in the **Related Entity ID** column as ID 12; which consists of 1 record from the employees data source.

From the same screen you can compare the employee data source to any other loaded.



1. Change the B data source from None to SalesForce.com Customers.
2. Notice how the circles become Venn diagrams to show the overlap between the A and B data sources. Observe there is 1 employee duplicate, 60 SalesForce.com duplicates and 2 employees that are also SalesForce.com customers.
3. Click the number 31 in the **Possibly Related** diagram.
4. Senzing resolves records into entities, notice how **Entity ID** 426 consists of two SalesForce records (James/Jimmy Foster) who share an address and phone with the **Related Entity ID** 2361 which consists of a single SalesForce.com customer (Jade Vaughn).
5. Consider entity 472 – a single SalesForce.com customer named Leonard Harrison who possibly lives with (shares an address and phone) with entity 2183 – Tom McKenzie – also a SalesForce.com customer record.
6. You can click the blue download icon on the far right to export the currently displayed matches to a CSV file for further review and processing.

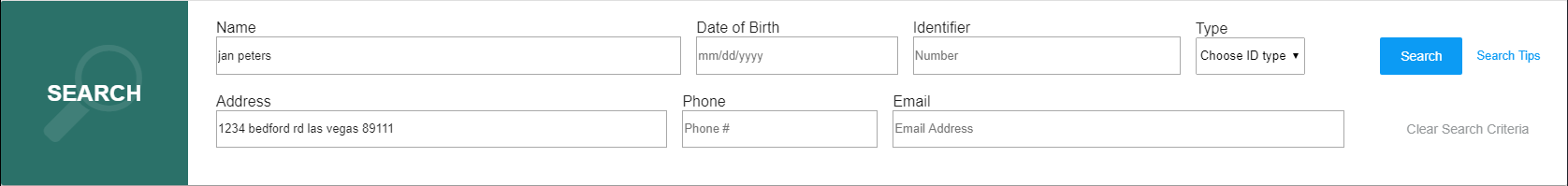
# Step 6: Review the Dashboard



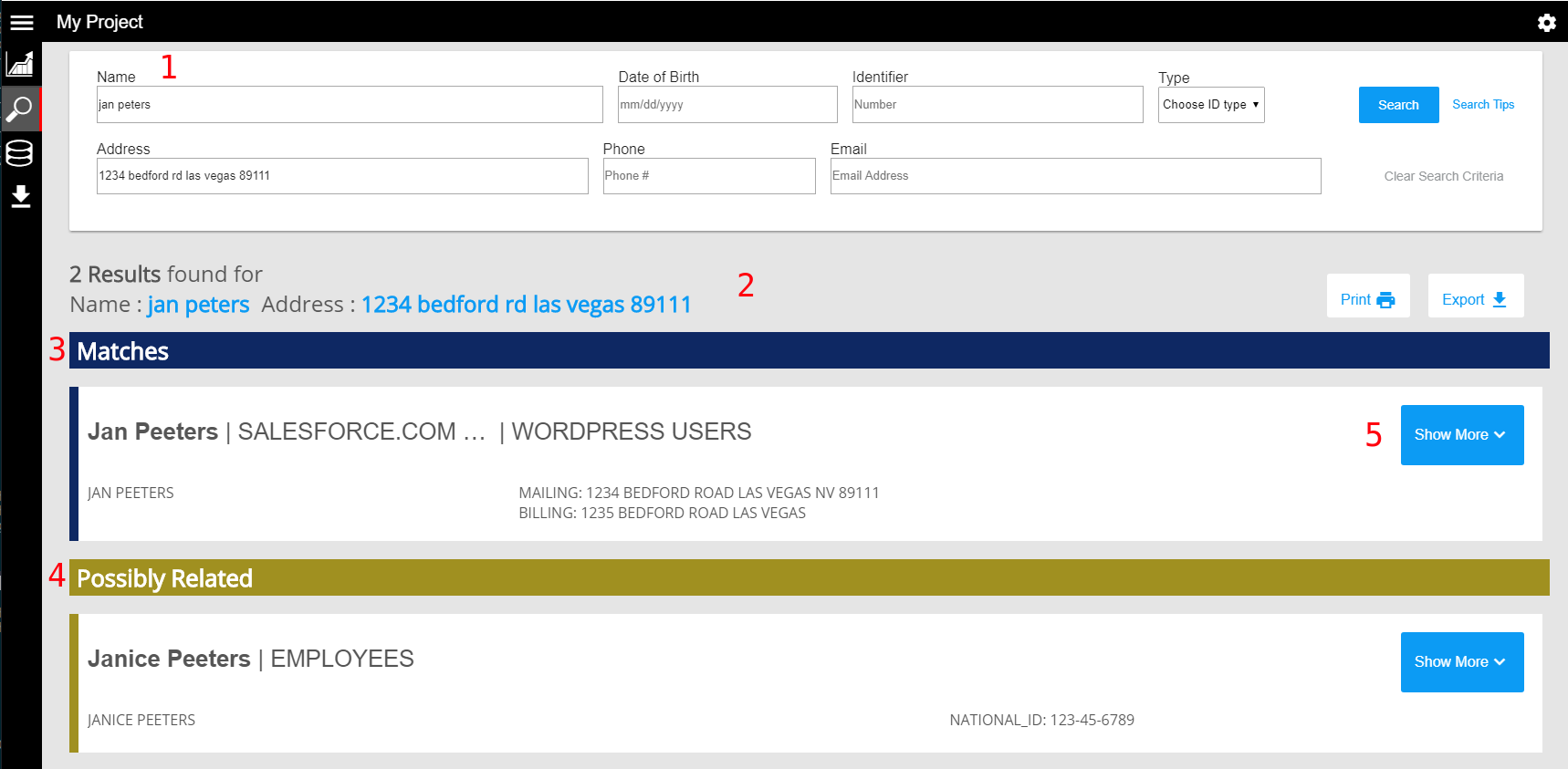
1. The Data card shows the number of records in each source on the left.
2. The Data card also shows how close you are to your licensing limits on the right and provides the ability to upgrade your limits with the **Upgrade License** button.
3. The Search card is for executing a single subject search.
4. The Review card is another place where you can select data sources to compare and review the matches within and across them.

# Step 7: Perform Single Subject Searches

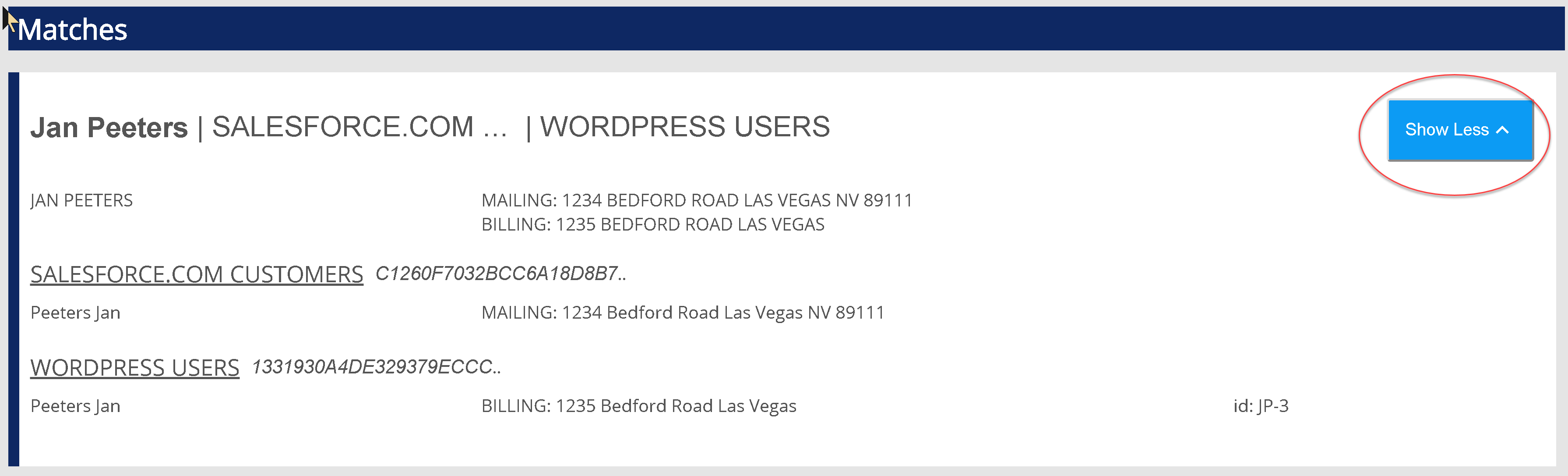
A single subject search is a search for a specific person or company. It’s not a query for everyone with the last name Peters or all Peters that live in Las Vegas.



This search for a Jan Peters at 1234 Bedford Road, Las Vegas, NV 89111 brings up the following screen:



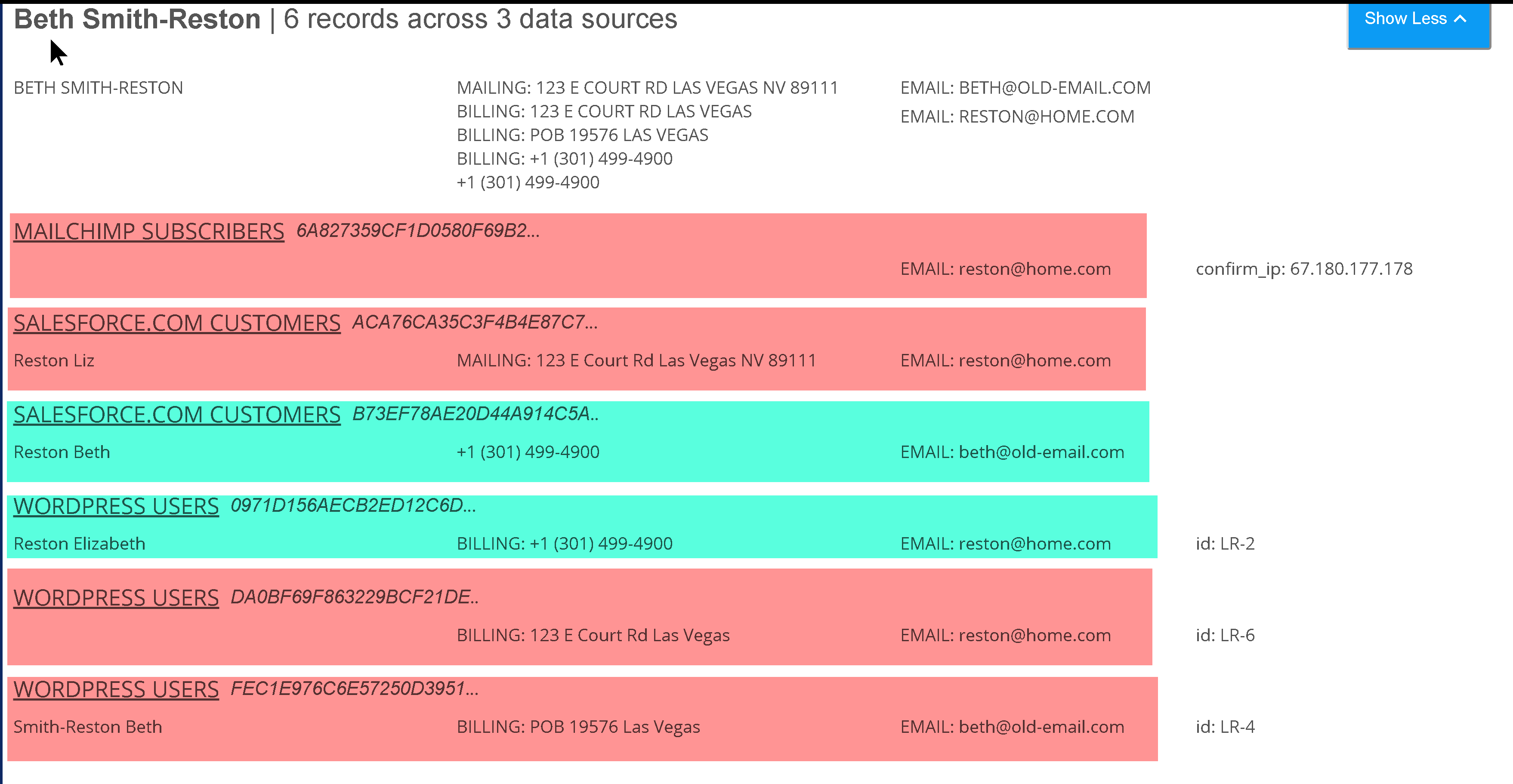
1. The screen maintains the search form at the top, you can execute further searches without returning to the dashboard. You can also get to this screen via the magnifying glass on the left-hand side.
2. The result summary shows your search criteria, how many entities were returned and has buttons for printing and exporting the search results to a CSV file for further analysis and review. Senzing returns all matches for the search criteria grouped by Matches, Possible Matches, and Possibly Related records.
3. In this case, there is a **Match** to a Jan Peeters entity on name and address despite the badly spelled last name. Note, we have captured variations to their address and a search for either 1234 or 1235 Bedford in Las Vegas would return the match.
4. There is also a **Possibly Related** match to an employee entity - Janice Peeters - despite name variations in both first and last name. The ability to match names through variations such as these is an extremely important feature of our technology. You don’t have to rely on the person searching to know and try all the name variations, synonyms, alias, etc. We do this automatically for multi-cultural names across the globe.
5. Remember, Senzing resolves records into entities. The **Show More/Less** button on each entity returned will show or hide the records each entity consists of so you can see which record contained which values. This is the resume for the identified entity.



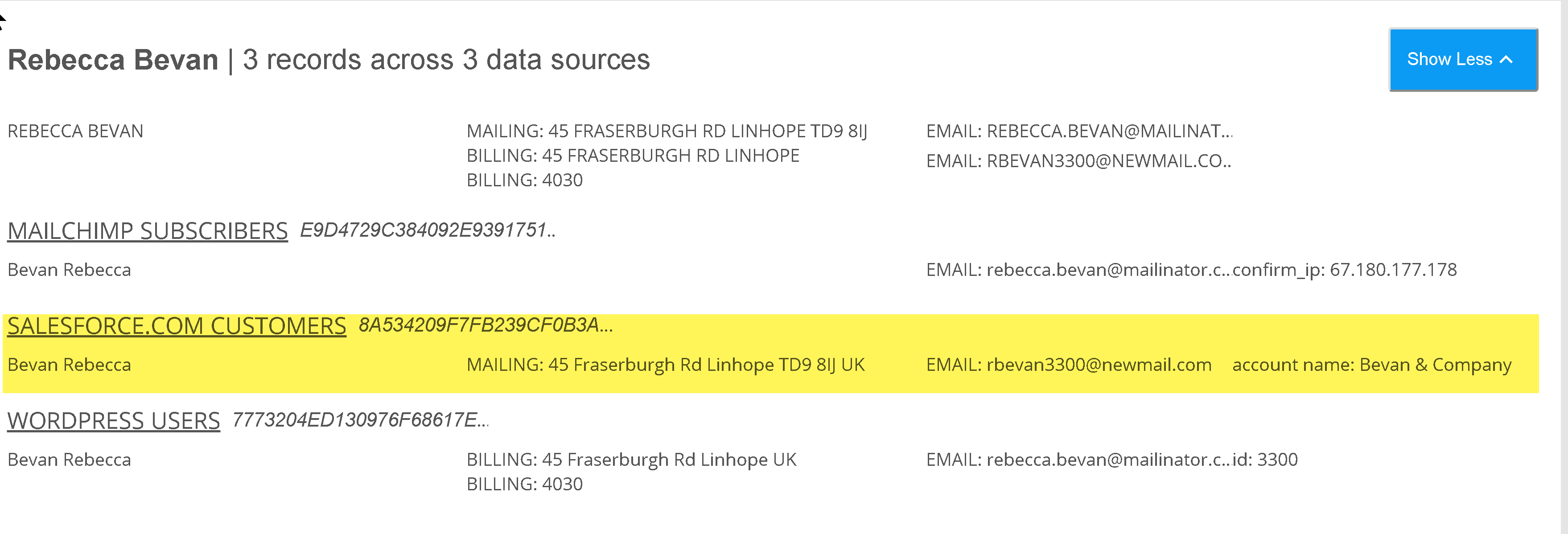
Additional searches to try:

* **Name**: Liz Reston  
  **Phone**: 301-490-4900

Notice only two of her six records contain the name and phone number searched for. This is entity centric learning at work. We resolve the entities during loading so later searches find the complete entity using a single search.

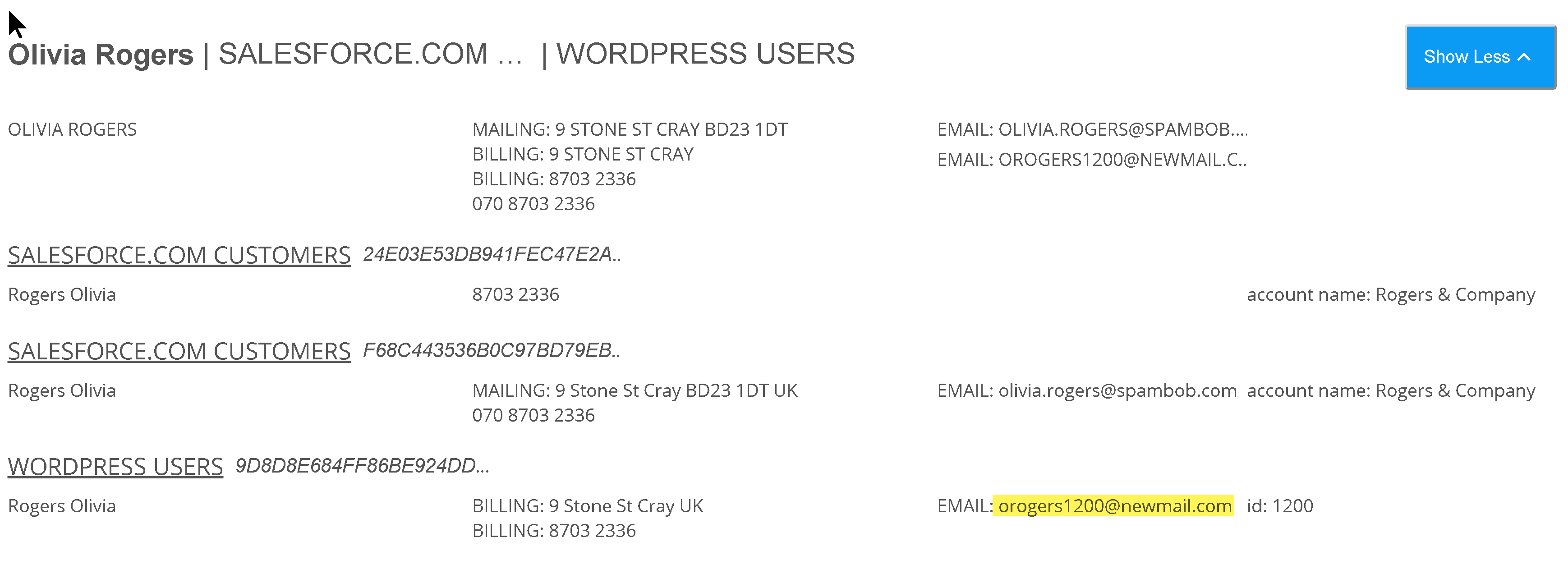


* **Name**: Becky Bevan   
  **Address**: 45 FRASERBURGH RD LINHOPE



What is the most complete address and reliable email for Rebecca? Consider Salesforce to be the most trusted data source given these three data sources.

* **Email**: orogers1200@newmail.com



Would you know Olivia was also in your Salesforce.com database?

# Next Steps

We hope you have found Senzing easy to use and get started with, providing a unique and powerful approach to resolving entities across and within data sources. Your next step is to load your own data sources.

Be sure to check out the getting started information on the initial screen upon opening Senzing. Feel free to contact us should you have any questions.