

EASY START GUIDE IMPLEMENTER

Version 14.7

Table of contents

About this tutorial	4
Using streamed data or data from other databases	4
Objectives and outcomes	5
Installing Siren Platform - Easy Start	6
Installing and launching Siren Investigate	6
Installing on a Windows machine	6
Installing on a Mac or Linux machine	7
Welcome to Siren Investigate	9
Importing data into Siren Platform	10
Downloading the data	10
Importing your first table of demo data	11
Defining field data types	11
Setting the properties of the entity table	12
Terminology	12
Creating dashboards automatically	14
Auto-generating a dashboard	14
Editing dashboards	16
Using the dashboard	17
Creating filters	17
Performing a text search	18
Creating dashboards manually	19
Creating the investments entity table	19
Creating the dashboard manually	20
Creating visualizations	20
Setting the dashboard data model	23
Importing the investors table	24
Relations in the data model	24
Creating relations between entities	26
Creating relations manually	26
Visualizing the data model as a graph	27
Creating a self-relation	27
Dashboard-to-dashboard associative navigation	29
Creating a dashboard for investors	29
Adding the Relational Navigator to the other dashboards	31
Recap: Where are we now?	31

Exercise: Investigating investments	33
Part 1: Filtering a dashboard	33
Part 2: Investigate the relation of data between two entity tables	35
Part 3: Digging deeper with relational navigation	36
Part 4: Exploring results with the link analysis in the Graph Browser	37
The Graph Browser: Link analysis	42
Embedding the Graph Browser in dashboards	42
Using the Graph Browser	43
Advanced features	46
Creating entity identifiers (EIDs)	47
Creating a 360 dashboard	48
Configuring the Dashboard 360	49
Filtering with Dashboard 360	51
Importing and transforming data	52
Transforming your data	53
Configuring the Graph Browser	57
Fixing labels with graph lenses	57
Natural Language Processing (NLP)	58
Creating an NLP-based Record Table	60
Connecting extracted entities in the data model	61
Example usage of NLP Data in Siren Investigate	62
Creating a Topic Explorer visualization	65
Deleting data and changing a table schema	66
Legal notices	67

About this tutorial

You can use Siren Platform in different kinds of scenarios, varying the way you import data depending on the data type.

This tutorial helps you to get started with the <u>Siren Platform</u> version 14.7 or later. It helps you to install Siren Investigate as an empty platform, import demo data, and start exploring and analyzing the data. It focuses on using static data, which we import from Excel or CSV files. Static data, which does not change or changes only a few times a day, is commonly used in Business Intelligence (BI) or broader knowledge discovery scenarios.

Using streamed data or data from other databases

For use cases that involve streaming data, such as in **cybersecurity**, **operational intelligence**, **log management**, or **IoT**, data streams can be loaded by external applications that write directly to the underlying Elasticsearch cluster. Examples of external tools for these scenarios are <u>Logstash</u>, <u>Beats</u>, <u>Fluentd</u>, or <u>Streamset</u>. It is also easy to write data to Elasticsearch directly by using the APIs.

Siren Platform also supports working on data in remote JDBC data sources. This can happen directly, with no data copy, or by using the built-in, UI-assisted, *reflection* process, where data is copied and periodically refreshed. Refer to our <u>documentation</u> to understand how to use data streams and JDBC datasources.

Objectives and outcomes

This tutorial covers the following tasks:

- 1. Installing Siren Platform.
- 2. Importing CSV files.
- 3. Creating an entity table.
- 4. Creating a dashboard both automatically and manually.
- 5. Adding interactive visual filters and textual search to the dashboard.
- 6. Creating an associative data model.
- 7. Creating a relational navigator visualization to navigate dashboard-to-dashboard.
- 8. Using Graph Browser and Link Analysis including test exercises.

The tutorial also includes an <u>Advanced features</u> section that contains additional important topics.

By the end of the first part of this tutorial, you will be able to use Siren Platform to search and investigate data about companies, investments, and investors. You will learn how to apply basic search and analysis functions to the data.

If you go on to complete the exercises in the <u>Advanced features</u> section, you will discover how to import and investigate data using some of our advanced features and configurations.

Check out our <u>What is Investigative Intelligence</u> video for insight into what's possible.

Installing Siren Platform - Easy Start

Operating system	Supported browsers	RAM
Windows*	Google Chrome	8G
macOS	Mozilla Firefox	
Linux		
*Windows machines need an improv ZIP support cannot handle the distrik	ved ZIP manager such as <u>7-zip</u> or <u>WinF</u> pution's large number of files.	RAR because the standard Windows

Confirm you meet the following system requirements:

Installing and launching Siren Investigate

To download Siren Investigate, complete the following steps:

- Go to <u>http://www.siren.io/downloads</u> and download Siren Platform Easy start.
- 2. Save the compressed file to a local folder named **siren-platform**.
- 3. Go to <u>www.siren.io/trial</u> and request a community or trial license.

Installing on a Windows machine

- 1. Extract the compressed file. It contains two folders:
 - **elasticsearch**: Contains the self-contained, single-node Siren Enhanced Elasticsearch cluster.
 - **siren-investigate**: Contains the Siren frontend application accessed through a Web browser.
- 2. Double-click **start.bat** and wait for both Elasticsearch and Siren Investigate to load.
- 3. When the installation is complete, in a browser open http://localhost:5606.

NOTE Do not close either the Elasticsearch or Siren Investigate command windows until you are finished working with Siren Investigate.

IMPORTANT The first time you run the installation, the Investigate window may appear to be stuck. This is the optimization phase. It continues writing log messages after a few minutes.

Relaunching on Windows

To relaunch Investigate, repeat steps 2 and 3 of the above procedure.

Installing on a Mac or Linux machine

NOTE Mac users: If you are running Catalina version 10.15 or later, do not extract the downloaded ZIP file yet. You must first remove an extended file attribute (xattr) before you extract the downloaded ZIP file.

NOTE Linux users: Extract the downloaded file and go to step 6.

- 1. Rename the downloaded ZIP file siren-platform-*-darwin-x86_64.zip.
- 2. Download Amazon Corretto 17 for macOS from

https://docs.aws.amazon.com/corretto/latest/corretto-17-ug/downloads-li st.html

- 3. Open a **Terminal** window at the **siren-platform** location. For example, right-click the **siren-platform** folder and select **New Terminal at folder**.
- 4. Execute the following command:

xattr -d com.apple.quarantine siren-platform-*-darwin-x86_64.zip

5. From siren-platform, extract siren-platform-*-darwin-x86_64.zip

The extracted folder contains two folders:

- **elasticsearch**: Contains the self-contained, single-node Siren Enhanced Elasticsearch cluster.
- **siren-investigate**: Contains the Siren frontend application accessed through a Web browser.
- 6. Open the **elasticsearch** folder.
- Open a Terminal window at the bin folder and execute the following command: ./investigate

8. When the log prints

[console] Template [template:kibi-html-angular] successfully loaded in a browser, go to http://localhost:5606.

NOTE **Do not** close either the Elasticsearch or Siren Investigate Terminal windows until you are finished working with Siren Investigate.

Relaunching on a Mac or Linux

To relaunch, repeat steps 3 and 4 of the above procedure.

Welcome to Siren Investigate

Siren Investigate opens to Siren Search by default. When you have data uploaded to Siren Investigate you can immediately perform a search by a search term.

To expand the navigation menu and display the range of apps, click the sidebar on the left side of the screen.

	Siceo	Q Search all data	📾 All Articles (646,903) 🖓		2 Share Clone Edit Expo	rt Watcher Reset 🕓 Last 150y
0	JICII	A 🖶 Create	Filters Search in dashboard		2	Help Q
q	Search		Relational Navigator	Article Multichart Articles by Source	Topic Clustering - Articles	
Ĥ	Dashboard	All Articles 646.9k	mentions (18507 All Companies) →		Moreadt gaze	
ø	Discover	COMPANIES V	mentions (53446 Person) V	200,000 -	operating missed too	Android
l.l	Visualize	INVESTMENT ROUNDS			Other Topics	upuar
		💄 INVESTORS 🗸 🗸			10pics	
	Timelion	🗶 TOOLS 🗸 🗸			total)	
**	Access control	¬→ MY GRAPHS		Published year: Accending	ip aperating ip	ohon appl
*	Data import					o.p.p.
Ī	Siren Alert				market	
æ	Web services		Organizations Mentioned Persons Mentioned		percent wireless	inter appl
	Develop		os Motorola		so	cial and the second
	Devicous			Oracle	sales	orking
:C	Data model		l lech News a	and Analysis	network	Coogle
۵	Management		Dell AT&T N 4: and a		200	Google
			Reuters Search	SOTT Bead Sprint ACK	billion year Facel	book
			Sun Microsystems Hewlett-Pa	ckard App Store Congress	ago month un un un	1000 Au
			Sansong	PG5	Docs: 646.9k Sampled Docs: 100k	× ^×
			Articles NLP Articles table			-
			□ Time - Title	Snippet	Organizations mentioned	Persons mentioned
			07/01/2016 Seven perfectly legal ways to get Windows 7	cheap (or Update, September 2, 2010: Several of the deals listed in	this post are no longer available. For the	
HOME	Default Dataspace		even free)	latest on the Family Pack, be sure to read this post: Wind	lows 7 Family Pack discounts return.	
4	sirenadmin		28/07/2014 8 mobile apps for managing business cards	When was the last time you handed out or received a ph	ysical business card? As the popularity of	
	Log out			of them being printed and circulated. But old habits die l	and, and there's still something very	
ര	Нер			personal and deliberate about exchanging something ta	ngible after a particularly positive meeting.	
		Filter deckloseds by serve	Export 🛓 Add New 🕂			1-50 of 646,903 < >
0	Collapse	Filter dashboards by name				

Figure 1: Siren Investigate open on a dashboard

Callout	Description
1	App menu
2	Option menu

Importing data into Siren Platform

(Average completion time: 10 mins)

In this tutorial we use sample data from a TechCrunch data source that was collected some years ago and a sample of technical articles from the web.

Downloading the data

- 1. Download the sample data file.
- 2. Extract the compressed folder that contains the following files:
 - **companies.csv**: A list of companies that includes geo-locations and descriptions.
 - **investments.csv**: An associative table that connects companies to investors with amounts, investments dates, and round code: seed round or round A.
 - investors.csv: A list of investors.
 - **articles_nlp.csv**: A collection of technical articles. Most of the articles mention one or more companies. To extract these mentions, follow the advanced NLP section of this tutorial.
 - **additional_companies.csv:** A list of companies to use with the advanced feature <u>Importing and transformation</u>.

The main files are relationally connected. The following entity-relationship diagram shows the connections between the tables. There are also other relationships in the data, for example, city and state names, URLs, and email addresses.



Importing your first table of demo data

(Average completion time: 5 mins)

For your first sample dataset, import **companies.csv.**

- 1. Go to the **Data model** app and click **Create an entity table**.
- 2. Click Quick import from a spreadsheet.
- 3. Upload **companies.csv**.
- 4. Click Add fields.
- 5. In Table name, enter companies.

Defining field data types

Most of the data types that are listed in the **Type** dropdown menu, such as Integer and Date, are easy to understand. However, the following options require explanation:

Text

Use this option for long texts, such as emails and messages. When you choose **Text**, the backend system splits the content into individual words and calculates statistics on those words. This allows you to have word clouds and the best level of searchability.

Keywords

Use this option for unbreakable strings. For example, 'New York' and 'Los Angeles' are cities that must be marked as Keyword to avoid the individual words being treated as single-word search terms, which would then lead to wrong results later.

Advanced Mapping

The **Advanced Mapping** switch gives advanced users the option to add a JSON object to

define the mapping in Elasticsearch format. For example, selecting **Date** and switching on

Advanced Mapping allows you to define a specific format for the date field, such as

{"format":"yyyy-MM-dd HH:mm"}.

Set the data types for the following fields:

- 1. Scroll to description and from Type, select Text (allow word cloud).
- 2. For **founded_date**, set the type to **Date**.
- 3. For **Geopoint**, set the type to **Geo Point**.
- 4. For **overview**, set the type to **Text (allows word cloud)**.
- 3. Click **Next**.
- 4. On the **Transform data** screen, click **Next**. You don't need to do any data transformation now.
- 5. On the Load data screen, click Start loading.

When the import is complete, the details of the new entity table are displayed.

If you already have data in Elasticsearch, you can make it visible in Siren by creating an entity table from an existing index pattern. For more information, refer to the <u>documentation</u>.

Setting the properties of the entity table

- 1. Go to the **Data model** app and click the **Info** tab.
- 2. In **Icon**, select an icon to represent companies, such as a building like **fal fa-building**.
- 3. In **Color**, choose a color for the icon.
- 4. From Default label, select Document Field then label.
- 5. In **Time filter**, click **Enable the Time filter based on the following field:** and select **founded_date**.
- 6. From the options menu in the upper-right corner, click **Save**.

Terminology

- **Record:** In this case, we have created one record for each row in the CSV file. Records contain fields. Fields contain values.
- Entity table: A collection of records that contain information about the same type of entity, such as the companies table. An entity table refers to one or more underlying indices. The term index is used by Elasticsearch, Siren's backend system. Entity tables are the core 'searchable objects' in the Siren Platform. This means that they can be searched, drilled down, and become the basis of analytical visualizations.

- Index pattern: Entity tables composed of more than one indices are called an index pattern. For example, entering logs_* as the index pattern will create an entity table that displays records coming from all the Elasticsearch indices that have a name starting with logs_ (for example, logs_2021_10_1, logs_2021_10_2).
- **Searches:** Typically, searches are a subset of an entity table. For example, if you wanted to sort your list of companies by those based in France, you can create a search for French companies, which would then appear as a subset of the companies entity table.

Creating dashboards automatically

The Siren Investigate dashboards display a set of visualizations in a customizable grid layout. Each visualization is normally linked to a search, such as companies or investments. You can organize dashboards into dashboard groups, shared, and customized.

You can create dashboards both manually and automatically. To manually create a dashboard, you need to create and add each visualization to your dashboard. Alternatively, the dashboard generator can automatically create the visualizations. This is a good way to get started on a new dataset.

After you create an Index Pattern Search, the following message appears:

No dashboard exists for this index pattern search
 To generate a dashboard, either select the fields you want to include or click Discover most relevant. Then, click Generate a dashboard.

Auto-generating a dashboard

(Average completion time: 5 mins)

You can access the auto-generate dashboard feature through:

- The **Data** tab of the **Data model** app.
- The **Discover** app.

Generate a dashboard for the companies data as follows:

- 1. In the **Data model** app, click the **Data** tab to view the records in the entity table. To the left there is a list of **Available Fields.**
- 2. Click **Discover most relevant fields**. The system identifies the fields that are most relevant for generating the dashboard.
- 3. Accept the default fields and click **Select.** The selected fields are listed on the left.
- 4. Click Generate a dashboard.
- 5. In the **Generate a prepopulated dashboard** window, in **Title**, enter Companies and click **Generate**.
- 6. In the Generate a prepopulated dashboard Report screen, click Create the dashboard.

The **Generate a prepopulated dashboard - Report** screen contains the following elements, however, we do not need to adjust them now. ① - Checkboxes that allow you to individually select which of the auto-generated visualizations to use in the new dashboard.

 2 - Dropdown menus, where you can select the best type of visualization (Chart Type) for each field.

③ - Text boxes that allow you to personalize the name of each visualization (Vis Name).

~	Type 🗘	Field ≑	Chart Type 🗘	Vis Name 😂
~	@	Geopoint	Enhanced Coordinate Map	Geopoint locations map
/	t	category_code	Vertical Bar	Top 20 category_code
2	t	city	Analytic Table 🔹	List of city
/	t	countrycode	Analytic Table 🔹	List of countrycode
4	t	description	Tag Cloud 👻	Top 20 description
/	0	founded_date	Line	Documents count by founded_date
2	t	hasstatus	Pie 🔹	Top 20 hasstatus
2	#	number_of_employees	Vertical Bar 🔹	Histogram of number_of_employees
-	t	overview	Tag Cloud 👻	Top 20 overview
~	t	statecode	Analytic Table 🛛 👻	List of statecode

Figure 4: Generate a prepopulated dashboard

The new **Companies** dashboard is now generated with different types of visualizations including an enhanced coordinate map, vertical bar charts, a pie chart, a line chart, analytic tables, word clouds, and a record table. For more information about visualizations, refer to the <u>Siren Platform documentation</u>.

If you don't specify a new dashboard name, it defaults to the name of the index pattern search. To edit the dashboard name, in the dashboard menu, right-click the dashboard and select **Rename**.

TIP

Editing dashboards

(Average completion time: 5 mins)

You can organize the visualizations on the dashboard. You can resize, move, and stack the visualizations in tabs.



Figure 5: Visualizations layered in tabs

- 1. From the option menu, click Edit.
- 2. Drag the visualizations to different locations on the dashboard or stack the visualizations.
- 3. Resize visualizations that contain graphs and charts to see the visualization.
- 4. Click Save > Save dashboard.

Using the dashboard

Explore this dashboard and interact with the visualizations to take a closer look at the data.

Creating filters

(Average completion time: 5 mins)

You can create filters for dashboards by either using the **Add a filter** button or by selecting categories in the visualizations. Create filters by interacting with the visualizations as follows:

1. In the **Top 20 hasstatus** pie chart, click the funded status segment to filter the dashboard on that category. A filter appears at the top of the dashboard and the number of records in the dashboard is reduced.



Figure 6: Filter indicated on dashboard

- On histogram visualizations, you can create multiple types of filters. On the Top 20 category_code bar chart, press and hold Ctrl/Cmd, and click the software and web category codes bars.
- 3. Click **Apply Now** (1) to apply a filter with an OR condition.



Figure 7: Creating multiple filters

Performing a text search

You can use the dashboard search bar to perform a free text search. Before you perform the text search remove all the current filters, hover over each filter and click the remove icon (trash).



1. In the dashboard search bar, enter **cloud**. The dashboard updates all of the visualizations to reflect the search results.





- 2. Perform an advanced search by using the Lucene query syntax.
 - To search for a value in a specific field, prefix the value with the name of the field. For example, enter category_code:software to find all records that have software as the category code.
 - Use the Boolean operators AND, OR, and NOT for complex queries. For example, enter category_code:software OR category_code:web to get all of the results with either the software or web category code.
 - You can also use the **fuzzy operator** '~' to search for terms that are similar to, but not exactly matching the search terms. For example, type **OR commerce~** to get the results of all mentions of ecommerce, commercial, and so on.
- TIP
 To reset the search bar to the previous entry cloud, click **Reset** in the Options menu in the upper right of the screen.
 To reset the dashboard, and all dashboards at once, to its previous saved state, click the **Reset all** filters button in the Dashboard sidebar.

Creating dashboards manually

Import another table and manually create the dashboard.

Creating the investments entity table

(Average completion time: 5 mins)

Create a new entity table to use to create a new dashboard. Import the spreadsheet, define the data type, and set the properties of the entity table like you did in <u>the first import</u>.

- 1. Go to the **Data model** app and from **Entity tables**, click the Add/Create button.
- 2. Import investments.csv and name the entity table investments.
- 3. Under Field, scroll to funded_date and set the type to Date.
- 4. For **raised_amount**, set the type to **Long.** This allows an integer over 32 bits for large investments.
- 5. In the lower-right corner, click **Next**.
- 6. On the Transform data screen, click Next.
- 7. On the Load data screen, click Start loading.
- 8. Go to the **Info** tab.
- 9. Select an icon to represent currency, such as **fal fa-money-bill-alt** and select a color.
- 10. In Default label, select Document Field and raised_amount.
- 11. In **Time filter**, click **Enable the Time filter based on the following field:** and select **funded_date**.
- 12. In the upper-right corner, click **Save**.

Creating the dashboard manually

(Average completion time: 7 mins)

NOTE This is an optional task, you can use the auto-generate option and skip to Importing the investors table.

Create a dashboard manually for **investments** as follows:

- 1. Go to the **Dashboard** app.
- 2. Click Create > Dashboard.
- 3. Name the dashboard Investments and click Create.

At this point, you have created a dashboard, but it does not contain any visualizations and it is not linked to an entity table.

Creating visualizations

(Average completion time: 15 mins)

Add visualizations in the following order:

- 1. Vertical bar chart
- 2. Record table
- 3. Analytic table

Later, you will add the visualizations directly from the current dashboard. You can also add them by using the **Visualize** app.

Vertical Bar Chart

- 1. In the upper-right corner, click **Edit > Add > Add a new visualization**.
- 2. From the **Charts**, select **Vertical Bar**.
- 3. On the **Select an entity table or search** screen, select the **investments** entity table.
- 4. In the **buckets** pane, click **Add buckets** and select **X-Axis** ①.

Implementer Easy Start Tutorial - Siren Platform version 14.7



Figure 9: Adding buckets

- 5. From Aggregation, select Terms 2.
- 6. From **Field**, select **round_code** ③
- 7. In **Size**, set the value to **20** ④.
- 8. Click the **Apply changes** button (5) to see a preview of your chart.
- 9. From the options menu, click **Save** 6.
- 10. On the **Save** panel, name the visualization **Investment number by round** code and click **Save to dashboard**.
- 11. You are redirected to the **Investments** dashboard in edit mode, where you can resize or reorder the visualizations. Click **Save** > **Save dashboard**.

Record table

- 1. From the dashboard, click **Edit > Add > Add a new visualization**.
- 2. Select **Record Table**.
- 3. On the Select an entity table or search screen, select investments.
- 4. The default values are adequate, so click the **Apply changes** button.
- 5. Click **Save...**, name the visualization **Record table** and click **Save to dashboard**.
- 6. You are redirected to the **Investments** dashboard in edit mode, click **Save** > Save dashboard.

In a record table, you can remove the columns that you do not need by clicking X beside the column name. To move columns, use the << >> buttons.

Analytic table

The analytic table visualization is similar in appearance to the record table. The difference is that you can view individual records with a record table, while you can make summaries with an analytic table.

- 1. From the dashboard, click **Edit > Add > Add a new visualization**.
- 2. In **Textual**, select **Analytic Table.**
- 3. On the Select an entity table or search screen, select investments.
- 4. In the **buckets** pane, select **Split Rows**.
- 5. From Aggregation, select Terms.
- 6. From Field, select raised_currency_code.
- 7. From Order By, select metric: Count.
- 8. In Size, set the value to 10.
- 9. Click the **Apply changes** button.
- 10. From the options menu click **Save..**, name the visualization **Analytic table**, and click **Save to dashboard**.
- 11. In the dashboard screen, click **Save > Save dashboard**.

TIP

Setting the dashboard data model

(Average completion time: 3 mins)

Before going any further, you must specify a type of data model for the dashboards.

Investigate has three dashboard options:

- **Dashboard does not represent an entity table**: Use when the dashboard's main function is to provide summary information and it does not need to be linked to an entity table. These dashboards do not display a record count or an icon.
- **Dashboard represents an entity table:** Use when the dashboard represents a single specific entity table, enabling the following:
 - The dashboard displays the count and an icon associated with the entity table.
 - Components like the relational navigator work and can be used on the dashboard.
 - You can add the content of the dashboard to a Graph Browser for link analysis.
- **Dashboard 360 with filter strategy**: Uses a dashboard-specific data model, which enables a single dashboard to contain visualizations that are based on different entity tables, and perform coherent filtering across all of them. This type of dashboard depends on Siren's relational capabilities and the creation of a data model. This is covered later in the tutorial.

You have created a dashboard with visualizations linked to the **investments** entity table. Next, specify that this dashboard is dedicated to the **investments** entity table.

- On the Investments dashboard, from the options menu click Edit > Data Model.
- 2. Select **Dashboard represents an entity table** and select **investments** from the dropdown list.
- 3. Click Save > Save dashboard.

Importing the investors table

(Average completion time: 5 mins)

Go to the **Data model** app, use the <u>same procedure that you used previously</u> to import **investors.csv** and name the entity table **investors**.

- 1. From the lower-right corner, click **Next** to accept the default field mappings.
- 2. On the Transform data screen, click Next.
- 3. On the Load data screen, click Start loading.
- On the Info tab, select an icon to represent investors, such as fal fa-user-chart, and select a color.
- 5. In Default label, select Document Field and label.
- 6. In **Time filter**, click **Disable the Time filter in dashboards** because this is not a time-based entity table.
- 7. From the options menu, click **Save**.

The **Entity tables** section of the **Data model** app now contains three entity tables: **companies, investments**, and **investors**.

IMPORTANT Do not create a dashboard for investors until you define the relations in the data model.

Relations in the data model

The power of Siren Platform is in its associative data model; defining how your entity tables are interconnected by their relations. It enhances the navigation across related dashboards and drives the link analysis feature, the Graph Browser.

In Siren Platform, a relation is a labeled link between entity types. You can think of an index pattern search as an entity type, which means that two index pattern searches can be connected by a relation in the data model.

You create these relations in the UI. In the example below, the Investments table is connected to the Companies table as follows:

Implementer Easy Start Tutorial - Siren Platform version 14.7



Figure 10: "Secured by" relation: Investments are secured by companies

When you define a relation, typically you must specify both a direct and indirect passive relation. For example, the companies *secured* the investments and the investments were *secured by* the companies

To view the data model graph:

- Go to the **Data model** app, select the **companies** entity table and go to the Data model graph tab.
- 2. Click the **Toggle entity highlight** button (lightbulb) to show all entity tables clearly. You can see that the indexes are not connected.

Creating relations between entities

Define the relationship between the indexes and create a relational data model.

Creating relations manually

(Average completion time: 5 mins)

Create relations to connect the entity tables. Use the **investments** saved search to specify which fields of other searches relate to it. Connect the **investments.companies** field with the **companies.id** field; a unique value in the companies data, based on the <u>relational model of the data</u>.

Create the relation between investments and companies as follows:

- 1. In the Data Model app, click the investments entity table.
- 2. Go to the **Relations** tab and click **Create a relation**.
- 3. In the From section, from Field, select companies.
- 4. In **Direct**, enter **secured by** and in **Inverse**, enter **secured**.
- 5. In the **To** section, in **Search**, select **companies** and from **Field**, select **id**.
- 6. Click **Save relation**.

Create the relation to the investor as follows:

- 1. Click Create a relation.
- 2. In the From section, from Field, select investors.
- 3. In Direct, enter made by, in Inverse, enter made.
- 4. In the **To** section, in **Search**, select **investors** and from **Field**, select **id**.
- 5. Click Save relation.

Visualizing the data model as a graph

(Average completion time: 1 min)

To see all the relations in the data model at a glance, return to the **Data model** graph tab.

From this screen, you can interact with the graph:

- Drag the icons to rearrange them.
- Change the direction of the relation labels by clicking the **Toggle relation direction** button.
- Change the layout.

Creating a self-relation

(Average completion time: 2 mins)

A self-relation is a relation between two entities of the same type, for example, a person who is *friends with* other people or companies that are *competitors of* other companies. These are defined in the same way as entities of different types.

In our sample data, the company records have a field called **one_competitor** that lists the ID of the main competitor of a company. Create a self-relation from **companies.id** to **companies.one_competitor** as follows:

- Go to the **Data model graph** tab, double-click the **companies** icon to switch to the **companies** entity table.
- 2. Click the **Relations** tab and click **Create a relation**.

Note: The relation between companies and investments is already listed because we specified this relation through the investments entity table.

- 3. In the From section, from Field, select id.
- 4. Enter **competitor** for both relations.
- 5. In the To section, in Search, select companies and one_competitor.
- 6. Click Save relation.

Implementer Easy Start Tutorial - Siren Platform version 14.7



Figure 11: Investment node with relations to an investor and a company

Dashboard-to-dashboard associative

navigation

The **Relational Navigator** is a special visualization that automatically shows links to connected records in other dashboards. By clicking on these buttons you perform an *associative* navigation, going from one set of records to another set of connected records.

For these buttons to appear in a dashboard, you must add a **Relational Navigator** visualization, the same way you add all visualizations to dashboards.

NOTE You can create a single Relational Navigator visualization and reuse the same visualization across all dashboards. When it is added to a dashboard, it shows buttons that connect with other dashboards that are relationally connected.

The dashboard generation wizard adds this visualization only if you have defined the relations between the data. Otherwise, you can add it manually.

Creating a dashboard for investors

Since we have already defined the relations for investors, we can use the wizard to create a dashboard for it as follows:

- Go to the **Data model** app and select the **investors** entity table.
- Follow the steps to <u>auto-generate the dashboard</u>, name the dashboard investors and click Generate > Create the dashboard.

One of the visualizations on the **Investors** dashboard is the relational navigator visualization, called **Related dashboards**. This visualization has a link to the

Investments dashboard. You can use this link to find all of the investments that were made by the investors on the current dashboard.

Filters	Search (e.g. status:200 AND extension:PHP)		
Related d	lashboards	0	List of affiliati
<u>Made</u> (28	723 Investments)	\rightarrow	affiliation_r
			Unaffiliated
			Google
			Facebook
			LinkedIn

Figure 12: Relational link to the investments made

- Filtering the records on the dashboard automatically shows a filtered count on the relational navigator link. For example, if you click the 'person' section of the **Top 30 investortype** pie chart. The record count on the dashboard and the count of investments are reduced as a result.
- Click **Made (28723 investments)** to go to the **Investments** dashboard. Due to the relational filter that is set, you see only investments that are associated with investors of type 'people'.



Figure 13: Investments dashboard now filtered by the investors

Adding the Relational Navigator to the other

dashboards

Add the Relational Navigator to the companies and investments dashboards to

find more connected records. For each dashboard, do the following:

- 1. Go to the **Dashboard** app and select the dashboard.
- 2. Click Edit > Add.
- 3. Search Related dashboards and click to add.
- 4. Click Save > Save dashboard.
- TIP When you are adding visualizations to dashboards, it is easy to also save filters by mistake. Either remove the filters when you are saving a dashboard or select **Don't overwrite currently saved filter and text query** in the **Save** panel.

Recap: Where are we now?

If you successfully completed all of the previous steps in this tutorial, your data

model should look similar to this:



Figure 14: Data model

You should have three dashboards: companies ①, investors ② and investments ③. The number shown after each dashboard name represents the number of documents in the relative entity table. Each dashboard should contain a relational navigator visualization, called **Related dashboards** ④. You can rearrange the visualizations to bring **Related dashboard** to the top.



Figure 15: Relational navigator visualisation on each dashboard

If everything looks correct, you can explore the data by completing exercises on each dashboard and visualizing the results in the **Graph Browser**.

Exercise: Investigating investments

Analyze the data to answer the four questions in this exercise. In the tasks you:

- 1. Answer a basic question by using the **Investors** dashboard.
- 2. Answer a question that involves a relational navigation from **Investors** to **Investments**.
- 3. Navigate further to the **Companies** dashboard.
- 4. Use a link analysis to explore the results together as a single picture.

Part 1: Filtering a dashboard

Question: How many investors are from Germany?

Basic solution

- 1. On the **Investors** dashboard, click **Filters** (1) > **Add a filter** (2).
- 2. Select countrycode, is and DEU (the ISO code for Germany) ③.
- 3. Click **Save** ④.

The dashboard filters to show records of investors from Germany only and the number of records, shown in parentheses next to the dashboard name, is updated.



Figure 15: Filter selections

NOTE At any time, you can click **Reset** to restore the dashboard to its saved state or **Reset all** to restore the state of all dashboards.

Alternative solution

You can also answer this question by adding a **Region Map** visualization. A **Region Map** is a clickable map that allows users to create filters in the dashboard in an interactive fashion. Unlike the **Enhanced Coordinate Map** visualization that works using geopoints (latitudes and longitudes stored in the records), the Region Map works by matching field values, such as **countrycode:deu**, with shapes and their names.

To add a region map visualization, do the following:

- 1. Click Edit > Add > Add a visualization.
- 2. From the map category, select **Region Map**.
- 3. In the Select an entity table or search window, select investors.
- 4. In the Visualize window, in Select buckets type, click shape field.
- 5. From **Field**, select **countrycode**.
- 6. You want to see data from more than five countries, so set **Size** to a value of **35**.
- 7. Go to the **Options** tab.



Figure 16: Options tab

- 8. From Vector map, select World Countries ①.
- 9. From Join field, select Three letter abbreviation (2).
- 10. Click the **Apply changes** button ③.

- 11. Click **Save...**, name the visualization **Investors region map**, and click **Save to dashboard**.
- 12. Move the region map into the position you want and save the dashboard.
- 13. To create a filter for investors from Germany (DEU), click Germany on the map. The dashboard count changes accordingly.

Answer: 212

Part 2: Investigate the relation of data between two entity tables

Question: How much money did German investors invest between 2010 and 2012?

Solution

Now that you have only German investors in the dashboard, you can use the relational navigator to go to the **Investments** dashboard. The visualization already displays the number of target records: Out of all investments, 575 were made by German investors.

- In the relational navigator visualization, click Made (575 Investments). A relational filter is automatically created on the Investments dashboard.
- 2. To change the time span, from the options menu, click Last 100y.
- In the Time Range window, click Absolute and enter the time span from 2010-01-01 00:00:00.000 to 2012-01-01 00:00:00.000.
- 4. Click **Apply**.
- To sum the value of the investments, click Edit > Add > Add a
 visualization and select Metric.
- 6. From the Select an entity table or search menu, select Investments.
- 7. In the **metrics** pane, expand **Metric** ③.

8. From Aggregation, select Sum ④ and from Field, select

raised_amount (5).

- 9. Click the **Apply changes** button **6**.
- 10. Click Save... and save the visualization as Total Investment Amount (\$).
- 11. Move the new visualization into position and save the **Investments** dashboard.



Figure 17: Metric visualization settings

Answer: German stockholders invested \$ 670,301,000 between 2010 and 2012.

Part 3: Digging deeper with relational navigation

Question: Of these investments, how much was secured by American companies?

Solution

- 1. Open the **Companies** dashboard. The number of records (159,782) is shown next to the dashboard name.
- In the List of countrycode visualization, click the Filter for value (magnifying glass) button next to USA to create a filter for American companies.

In the Related dashboards visualization, secured (34 Investments) shows

the amount of investments that match. Notice that, from our previous exercise, the Investments dashboard is already filtered for investments by German investors only.

3. Click **secured (34 Investments)** to reveal the investments on the **Investments** dashboard.

The filters on Investments dashboard has two relational filters:

- **coming from (212) docs on Investors** for investments from German investors.
- **coming from (46,017) docs on: Companies** for investments secured by American companies.

The **Total Investment Amount (\$)** visualization shows the sum of the 34

investments made by 22 German investors to 31 American companies \Im .



Answer: \$225,790,000

Part 4: Exploring results with the link analysis in the Graph

Browser

Now that you have solved all of the questions, use the link analysis feature in the Graph Browser visualization to explore the results. Explore how the companies, investments, and investors that we found in our investigation are linked together and if there are some interesting common connections. Before doing so, apply the relational filter on the **Investments** dashboard to focus our research around German Investments, Investors, and American Companies:

- 1. On the **Investments** dashboard, click **secured by (31 Companies)** and take a moment to see who these companies are.
- When you are ready, return to the **Investments** dashboard and click made by (22 Investors) and explore the investors.

If everything is done correctly, your filtered dashboards should now display:

- Investors 22 (German stockholders who invested in American Companies).
- Companies 31 (all of them American).
- Investments 34 (made by German investors and secured by the 30 American companies).

CAUTION If your dashboards don't have the correct filters applied and a large number of nodes are added, you might experience problems with system performance.

In this section of the tutorial, we'll get to see graphs like these:



Figure 17: Graph Browser example graph

Narrow your investigation on the Graph Browser

- 1. In the dashboard menu, expand **GRAPH DASHBOARD**, and click **Graph Browser**.
- 2. Drag the **Companies**, **Investments**, and **Investors** dashboards one-by-one into the Graph Browser.
- 3. From the **Layout** section of the toolbar, click **Standard** to rearrange the nodes.

Observe that two German stockholders invested in a company named Crowdpark. Explore this network further.



4. Set the cursor to selection mode and draw a rectangle around the network to highlight it.

Note: To change the cursor mode, from the left side of the Graph Browser, click the cursor button.

- From the toolbar, click Crop. This removes everything else from the Graph Browser.
- Before expanding this network, save this initial state. From the toolbar, click
 Save and give the graph a suitable name, such as "Shared investment".

Investigating the network

Now that you have saved your initial network, you can start asking more interesting questions. Before you proceed, take a moment to consider the links that might be interesting to explore.

We selected this network because the two German investors share an American company in their investment campaign. It would be interesting to uncover if:

- They have more common investments.
- They invested in CrowdPark in the same time period.

Question: Do the two German investors share additional companies in their

funding campaign?

- 1. While in selection mode, hold the **Ctrl**/ \mathbb{H} key on the keyboard and click each of the two investor nodes.
- When both nodes are selected, right-click one of them and select Expand by relation.
- 3. Select the made (27) investments check box and click Expand.
- Investigate whether one of the new investments that appeared is paid to a new common company. To do this, right-click any one of the investment nodes click Select > Select - By Entity Type.
- 5. In the pop-up window, select the **Investments** check box and click **Ok**.
- 6. All of the investment nodes should be selected. Right-click one of the nodes and select **Expand by relation**.
- In the pop-up window, select the secured by (27) companies check box and click Expand.

As you can see in the graph, the only common connection between the two German investors from 2010 to 2012 is CrowdPark.

Implementer Easy Start Tutorial - Siren Platform version 14.7



Figure 19: Crowdpark node with the two investors

Answer: No, the two German investors do not share additional companies in their funding campaign.

NOTE If you notice that the labels of the entities are not displayed correctly, ensure that you completed all of the steps when creating the relations.

Question: Did the two German investors fund CrowdPark in the same time period?

- Remove the nodes and return the graph to its saved state- click **Delete** > Delete All.
- 2. Click Add > Open saved graph > Shared investments.
- 3. In the **Open saved graph** window, click **Replace the current graph with the saved graph** and click **Add nodes**.
- Activate the timeline mode functionality of the graph- from the toolbar, click **Time**.
- 5. In timeline mode, hold the **Shift** key and select the relevant nodes; the company Crowdpark and the two investment nodes. This allows you to explore the temporal dynamics of the selected nodes.

As you can see in the following screenshot, Crowdpark was founded in the second quarter of 2009, while the two investments were made approximately one year apart.



Figure 20: Crowdpark and investors selected with timeline

Answer: No, the two German investors did not fund CrowdPark in the same time period?

To learn more about the graph, see the next chapter.

The Graph Browser: Link analysis

Embedding the Graph Browser in dashboards

Link analysis is one of the key features of Siren Investigate, it can be performed in the Graph Browser visualization.

In the previous exercise, you used a pre-configured dashboard with a **Graph Browser** component, however, you can add the Graph Browser to a dashboard as
a visualization by clicking **Edit** > **Add** > **Add a visualization** and selecting **Graph Browser**.

Using the Graph Browser

You can add data to the Graph Browser by dragging filtered or unfiltered data from dashboards. Data is represented as nodes.

You can add related nodes by clicking **Add nodes** on the toolbar and selecting **Expand by relation**. Reorganize the nodes using the **Layout** options.



Figure 20: Graph Browser toolbar

The Graph Browser also offers some advanced features to help with link analysis. The right-most sidebar contains the following tabs by default:

- **Styles** Create lenses to alter the way the graph looks. For example, create conditional formatting for labels, colors, sizes, or change visibility. For more information about lenses, see the <u>Siren Platform documentation</u> or follow the steps to create a lens in the <u>Configuring the Graph Browser</u> section of this tutorial.
- Relations Controls the way the graph expands, either when you click the expand button or when you double-click on nodes. Here, you can set filters by using Dashboard Filters or choose the data model relation to consider when expanding. For more information, see the <u>Siren Platform</u> documentation.
- Limits Applies the filters from a specific dashboard to the graph. This restricts the nodes that are produced from adding relations to a subset of those that are allowed by the current filters on a dashboard. For example,

select the Articles dashboard to count only the number of links to articles in your dataset. If you have filtered the Articles dashboard, then the filters are applied to the count.

- Cards- You can apply special views on the currently-selected nodes.
- **Selection** See details about the current selection and search within the selection. This feature is located in the bottom bar.

				Refresh counts
alations (11 of 34))		Ŧ	Relations
Relations - simp Q Search simp	ole (11/34)	•	0®	(11)
From V	Relation $$	To ~		Styles (9)
🗸 🔲 Articles	→ mentions	○ Location	0	
🗹 🥅 Articles	\rightarrow mentions	📱 Companie	000	Limits
📃 🔲 Articles	\rightarrow mentions	🗓 Companie	,e,	(0)
🗹 🔲 Articles	\rightarrow mentions	😵 Person	0	hil
🔽 🍭 City	\rightarrow office loca	💄 Investors	,e	Cards (6)
📄 🦁 City	ightarrow is headqua	👖 Companie	0	
🗌 🦁 City	\rightarrow is headqua	🗓 Companie	,e	
🛃 🞚 Companie	\rightarrow category	🗠 Company	0	
🗌 🞚 Companie	→ mentionec	🔲 Articles		
📃 🞚 Companie	\rightarrow secured	💷 Investmer		

Figure 21: Side panel

By default, the **Cards** tab is not displayed, you can add it by completing the following steps:

- 1. In the Graph Browser, in the upper-right corner, click Edit.
- 2. Click the pencil icon to open the Edit visualization menu.



Figure 22: Editing the Graph Browser

- Click the Scripts tab and, from the Cards section, click Add Card Script and select Histogram Card - default implementation from the dropdown menu.
- 4. Click **Apply changes** and **Save**. You might need to refresh the page for the changes to appear.

For more information about using the **Cards** tab, see the <u>Siren Platform</u> <u>documentation</u>.

Advanced features

This section of the tutorial teaches you how to use some of the more advanced features in Siren Platform.

- Creating entity identifiers (EIDs)
- Creating a 360 dashboard
- Importing and transforming data
- Configuring the Graph Browser
- Natural language processing (NLP)
- Creating a Topic Explorer visualization
- Deleting data and changing a table schema

Creating entity identifiers (EIDs)

Siren expands its analysis capabilities by introducing entity identifiers (EIDs). EIDs are used when entities exist (such as IP addresses, MD5s, or things that are identified simply by strings like City) that are present only as values in certain fields - in one or more tables - and do not have a dedicated table.

EIDs on the graph look like any other node, for example, an IP connecting two log records. They can also be used as 'intermediate concepts' when navigating dashboards.

To create an EID, complete the following steps:

- Go to the **Data model** app and click the **Configure an Entity Identifier** (plus) button.
- 2. In **Entity identifier name**, enter **city**, choose a suitable icon and color, and click **Create**.
- 3. Click the **Relations** tab.
- Create two relations; one with companies and one with investors, since both of them contain the city field. To manually create a relation, click Create a relation.
- 5. In Direct, enter office location and in Inverse, enter has office.
- 6. In **To**, select **companies** and **city**.
- 7. Click Save relation.
- 8. Repeat steps 4 to 7 for investors. In Labels, enter headquarter for and

headquartered in.

City (<i>En</i>	tity identi	ifier)					
Info	Data	Relations (2)	Data model graph				
① Create	a relation	Show inherit	ted relations			Expand all	Collapse all
From		R	elation		То		
🗌 🦁 City		\rightarrow	office location of	\leftarrow has office in	Investors		D
🗌 🦁 City		\rightarrow	is headquarter city of	\leftarrow has headquarter in	Companies		⊳

Figure 23: Relations

To see the EID added as a part of the data model, click the **Data model graph** tab.

Creating a 360 dashboard

The Dashboard 360 feature allows a single dashboard to contain visualizations that are based on different searches, and allows you to perform coherent filtering across all of them. Dashboard 360 uses the relational data model to enhance data analysis capabilities in the Siren Platform.

To create a 360 dashboard, first create a dashboard with two visualizations from different entity tables as follows:

- 1. Create a new dashboard called "All Companies".
- 2. Click Add > Add a new visualization.
- 3. Select **Record Table** and associate the record table with the **companies** entity table.
- 4. Click the Apply changes button and click Save....
- 5. Name the visualization "Table" and click **Save to dashboard**.
- 6. Click Add > Add a new visualization.
- 7. Select **Vertical Bar Chart** and associate it with the **investments** entity table.
- 8. In the **buckets** table window, click **X-Axis**.
- From Aggregation, select Histogram and from Field, select funded_year with the Interval set to 1.
- 10. Click the Apply changes button and click Save....
- 11. Name the visualization "Histogram of funded_year" and click **Save to** dashboard.

Configuring the Dashboard 360

Because two entity tables are involved, we must use a Dashboard 360 as follows:

- In the upper-right corner, click Edit > Data Model and select Dashboard 360 with filter strategy: Left filter join.
- 2. In the dialog box, select **companies** as the main search and click **Done**. The Mapping visualization to searches panel shows the list of Dashboard 360-compatible visualizations and whether or not they are assigned to a search node. The warning symbol indicates that neither of the visualizations created earlier are assigned yet.
- To assign a visualization, click the **companies** node to select it, then click the + button beside the required visualization (Table).



Figure 24: Main search node selected

- 4. You can now build the dashboard data model by defining how this search is to be relationally connected with other searches (that is, which path in the relational data model to take). To do this, click

 located on the lower right of the main search node. This opens the Select the path to destination window that lists all the possible relations that come from the chosen search, as well as all possible searches that can be reached through those relations.
- 5. In the **To** section, select **investments**, and click **Save**.

6. Select the investments node and from Unassigned visualization:, assign



Histogram of funded_year.

7. Save the dashboard.

The following screenshot shows how the dashboard looks with the visualizations assigned.



Figure 26: 360 dashboard

NOTE Your visualizations might be stacked as tabs. You can drag the visualizations to rearrange the layout.

Filtering with Dashboard 360

The benefit of Dashboard 360 is that you can perform coherent filtering across the dashboard's visualizations. For example, you want to see the pattern of yearly funding across a particular industry segment to investigate when funding for social media took off. To do this, filter the **Record Table** visualization on the category code **social**.

Þ	August 1st 2013, 01:00:00.000	Raploja	ecommerce
Þ	August 1st 2013, 01:00:00.000	Rootty	social QQ
Þ	August 1st 2013, 01:00:00.000	Epic Business	consul Filter for value

Figure 27: Filtering from the record table

The record table now lists all of the companies in the social category. More importantly, the histogram visualization displays the yearly funding trend; it shows a huge increase in the number of investments between 2012 and 2013.

You can also filter by clicking on the segments in the **Histogram of Funded Year** visualization; this will coherently filter the results in the Record Table.

Importing and transforming data

Typically, during investigation processes, investigators do not need or want to create an entity table. Instead, it is more common to add data to an existing source, possibly adapting the original data format to the target structure. In Siren, you can transform that data before you add it to an existing entity table.

Before you transform data, upload records from another CSV file as follows:

- 1. From the **Data model** app, click the **Add/Create** button ①.
- 2. Click Add data > Upload.



Figure 28: Add data

- 3. Upload **additional_companies.csv** from the sample data and click **Next**.
- 4. From **Target table** (5), select the **companies** entity table.

Data model				③ Last 100
Entity tables Companies Companies Entity Identifiers	© Extract data		2 Transform data	3 Load data
	Saved transformers None Vou can drag and drop the Source field in the	Save Site State	Save as to start creating a pipeline and transform	your data before the loading.
	Source table Q search field name or type	>>	6 Target table ^b None Q search field name ✓ None	Oreste a new table rimary key value? Fest transformer
	Source field Type Source sample		Field script	Target field Type Target output Target sample

Figure 29: Target table

Transforming your data

In the **Transform data** stage you can modify data before it is ingested. In this example, you are going to import some columns and merge them into the existing **companies** entity table. Before you proceed, have a closer look at the data transformation page after the last step.

You can drag the **Source field** ① into **Field script** ③ of the **Target field** ④. Knowing the content of **Source sample** ② and **Target sample** ⑥ makes it easy to merge the new CSV columns with the existing fields. More importantly, you can test the transform ⑦ before the import and preview the result in the **Target output** ⑤ column.

	Extra	act data		Transform data			3 Load data
Saved transfor	mers ~	🖺 Save 🔡 Save as					
You can drag and drop th	ne Source field in the	Field script box to start creating a	pipeline and	t transform your data before the loading.			Help and documentation
Source table	type		>>	C Search field name or type	Create a new table		X Define primary key value?
iource field	Type keyword	Source sample 2		Field script 3	Target field	Type keyword	Target output 5 linkedin.com/company/sch., http://www.500ftbags.com
ity	keyword	munich		S(category)	••• category_code	keyword	construction ecommerce
	knoword	Schoelder electric		Steind	ton, city	kmourard	and the later of t

Figure 30: Transforming data

Transforming fully compatible data

1. From the Source field section, drag category 1 into Field script 2 next to

the category_code Target field.

2. Click Test Transformer ③.



Figure 31: Category field

- 3. Repeat step 1 for each of the following compatible fields:
 - linkedin_address to blog_url
 - city to city
 - fundeddate to founded_date
 - website to url
 - label to label

Transforming indirectly compatible data

The target entity table **companies** has a Geopoint but the new CSV has two separate fields to denote location; latitude and longitude. The transformer makes this operation extremely simple:

- 1. Drag the lon field into the Field script of the Geopoint Target field.
- Edit the Geopoint field to accept the lat field also. In this case, add square brackets and the comma plus the space in the middle: [\${lon}, \${lat}].
 For information about formatting fields, refer to Elasticsearch

documentation.

3. Click Test transformer to check the result preview in the Target output.

Source table			>>	Target table	🗒 companies	~	① Create a new table
Q Search field name or t	ype			Q Search field name	e or type		
Source field	Туре	Source sample		Field script			Target field
label	keyword	Schneider-electric				4	••• email_address
lat	float	0.6365244177611469		{fundeddate}		4	••• founded_date
linkedin_address	keyword	linkedin.com/company/sch		[\${lon}, \${lat}]			••• Geopoint
lon	noat	0.8174781556149979				1	••• hasstatus

Figure 32: Transforming indirectly compatible data: lat and lon to geopoint

Sometimes, you need to combine values to match the target data format. This can be done with the concatenation operator **+** in **Field script**. For help with field scripts, refer to the <u>documentation</u>.

1. From **Source field**, drag **label** to the **Field script** of **id** ①.

label	keyword	Schneider-electric		4	••• email_address	keyword	50offbags@gm
lat	float	0.6365244177611469	\${fundeddate}	11	••• founded_date	date	2007-02-15
linkedin_address	keyword	linkedin.com/company/sch	[\$(lon), \$(lat)]	11	••• Geopoint	geo_point	
lon	float	0.8174781556149979		4	••• hasstatus	keyword	
region	keyword				••• homepage_url	keyword	http://www.50
website	keyword	Schneider-electric.com	"company/" + \${label}		•••• id	keyword	company/50of



 In Field script, enter the string "company/" followed by the + operator sign then \${label}.

NOTE: Ensure that your keyboard uses straight quotation marks. Curly quotation marks are not supported in this field and result in an error.

- 3. Click **Test transformer** and confirm that the **Target output** is similar to the **Source sample**.
- 4. Compare what you have done with the following screenshot to make sure you did not forget any fields and made all the necessary transformations.
- 5. In the lower-right corner, click **Transform data**. Some of the fields of the CSV were excluded to show that you can decide what to load and where.
- 6. Click **Start loading** to import the new CSV. In a production environment, any errors are listed here.

When the data is imported, you get a confirmation message. You can also check if the data imported correctly in the **Data** tab.

① Number of records before the import

② Number of records after the import

Data model								Sa			
Entity tables 🛛 😁	[©] 🔃 companie	es (<i>Entity</i>	table)								
siren-import-home-companies	Info Fields (38)	Data (159800) Relations	Scripted fields (0)	Options	Revisions	Search	🔆 Data model graph			
Entity Identifiers 🕒	Selected Fields	1 © б то ş	© No dashboard exists for this index pattern search To generate a dashboard, either select the fields you want to include or click Discover most relevant . Then, o								
	Discover mo relevant field	st ds Adv	arch d a filter +	4		Help	Q	Create child search C			
	Available Fields	۲	15.000	November 15th 1921, 1	7:29:40.777 -	November 15	th 2021, 17	29:40.777 — Auto			
Entity tables 🛛 😁	🛇 🔢 companie	es (<i>Entity</i>	<i>table</i>)								
siren-import-home-companies	Info Fields (38)	Data (262987	7) Relations	Scripted fields (0)	Options	Revisions	Search	i⊀ Data model graph			
Entity Identifiers	Selected Fields ? _source	2 C C To į	No dashboard generate a dashb	exists for this index p oard, either select the fie	a <mark>ttern sear</mark> o lds you want	to include or cl	ick Discove	r most relevant. Then, click			
	Discover mo	Se	arch			Help	Q	Create child search Cr			

Figure 33: Check Data tab to see the increased count

Configuring the Graph Browser

You can modify a graph in a number of ways by using the features in the Graph Browser.

Fixing labels with graph lenses

You can set graph node labels in the data model and then in the graph by using the lens feature. Always set them in the data model first. You can then override them in the graph with a lens that is useful for specific analysis. For example, you might want to see the article's title first, but change your mind and want to see the article's author as a label instead.

To create and activate a lens, complete the following steps:

- 1. Go to the **Dashboard** app.
- 2. In the sidebar, expand **Tools** and click **Graph Browser** or open a graph.
- 3. From the right-most side panel, click **Styles**.
- 4. From Create lens, select Label from expression 2.
- 5. In Lens name, enter Company labels ①.
- 6. From Entity Type, select companies 2.
- For Expression, click + and select label. It populates automatically with payload["label"].
- 8. Click **Save lens**.

Natural Language Processing (NLP)

(Average completion time: 40 mins)

The Siren Platform Easy Start version comes with the Siren Platform NLP plugin pre-installed. This adds an NLP pipeline processor to enrich documents that are being ingested in Elasticsearch, by a CSV file import or by another method, with entity annotations¹.

In this exercise, you use it to enrich another CSV file in our demo dataset, **articles_nlp.csv**, which contains tech articles that are harvested from the Web.

The pipeline processor takes a text field and produces annotations for Named Entities (Organization, Person, Location), which are then useful to connect to data that we already have. For example, connecting articles to companies or investors.

To do this, you activate a transformation pipeline with the NLP processor, while importing data into the **articles** index.

Before you begin

If you are using a slow machine, go to **Management** > **Advanced Settings** and edit the value for **ingest-xlsx:bulk_package_size** field to 20 and click **Save**.

Procedure

- Start importing the **articles_nlp.csv** file that is provided with the sample data. (Check the Importing data section if you need to refresh your memory.)
- 2. In the Table name field, type articles.
- Scroll down to the mapping definitions and set the **pdate** field as **Date** in the type column.
- 4. For the snippet field, select Text (allow word cloud) as the type.

NOTE This step is mandatory to allow the NLP processor to work.

¹ It is always possible in Siren Platform to use any other NLP engine as part of ETL, but Siren NLP is made available via the Elasticsearch pipeline processor, which makes it easy to activate and requires no external server or process.

5. Click **Next**.

6. On the Transform data screen, find the snippet field (1), click the ellipsis
(2), and activate Use Siren NLP plugin (3).

		1			 				3	
	Defines	structure			Transform data				Load data	
Saved trans	formore									
None	SIGITIELS	× B Save	Save:	as						
None I and			E Suve		C 1 1 1 C 11 1 1 1	P			0	
You can drag and	arop the source f	leid in the Field script box	to start crea	ting a pipeline and trans	form your data before the load	aing.			Q	Help and documentation
Source tab			~ ~	Target table	articlos		~			
	IC		>>	laiget table	al ticles					
Q Search field na	ame or type			Q Search field name	e or type			O × Define prima	iry key value? 🕜	D Test transformer
	_									
Source field	Туре	Source sample		Field script			Target field	Type T	arget output	Target sample
companies	keyword	company/sisso		\${author}	R		aution	keyword		
foundchrof	boolean	true		\$(companies)			foundchrof	boolean		
foundplorof	boolean	true		\$(foundeloref)			foundploraf	boolean		
id	keaword	data/techmeme/art		\$froundnipren			id	keyword		
image	keaword	data/techneme/art		\$(image)			image	keyword		
ndate	keyword	2007-01-10		\$(indate)			ndate	date		
protect	float	1		\$(pointe)			protect	float		
pyear	float	2007		\$(pvear)			pyear	float		
spinnet	keyword	2007		\$(spinpet)		2	X Multi-va	alued @		
source	keyword	Gigaom		\${source}						
title	keyword	iPhone Troubles: Ci		\${title}			Specify the del	limiter. For example, ',' or	'S-I	
url	keyword	http://gigaom.com/		\${url}			X Parse a	s JSON (2)		
	2			., ,		h	Use Sire	en NLP plugin @		
> Additional tra	insform pipeline	2								
									С	ancel Next >

The **nlp-pipeline** is now configured to perform the following tasks:

- Extract annotations by using various processors from the **snippet** field of the **articles** index.
- Write the annotations to the field **siren.nlp**.
- Extract the following items using regex or simple rules: Telephone, USTelephone, Email, IPv4, IPv6, MacAddress, Url, SortCode, HashTag.
- Extract Organization, Location, and Person Named Entities by using OpenNLP <u>NER models</u>.
- 7. Click Next and Start loading.

8. When the import process is complete, click the **Fields** tab and click the **Refresh field list** button. This allows Siren to discover the newly-found

entities.

ions	Search	🔀 Data moo	del graph	2
(r) S	ee Finger	print View	Add Fields	A Edit Mode

9. Next, auto generate a dashboard and name it Articles.

Creating an NLP-based Record Table

Create a special Record Table for tagging entities that were discovered by the NLP process in the processed text field (snippet).

- 1. In the Articles dashboard, in the upper-right corner, click Edit.
- 2. Click Add > Add a new visualization,
- 3. Select **Record Table** and select the **Articles** entity table.
- 4. In the **Cell formatters** pane, click **Add cell formatter** ①.

Visualize / New Visualization (unsaved)							Save Share Refresh
Add a filter 🕇							
Data Options	siren.importTimestamp	author	pdate	pmonth	pyear	source	snippet
Order and visibility	December 1st 2021, 11:26:09.642	•	January 10th	1	2,007	Gigaom	•
The order and visibility of columns can be set in the right panel; click on the [2] button to view all the available fields then click on the [1] button to togete the			2007, 01:00:00. 000				1
corresponding column.	December 1st 2021, 11:26:09.642	Sjvn	April 14th 2010, 02:00:00.	4	2,010	ITworld.c om	You think things are bad now between Apple and Adobe? Just wait until the lawsuit. 2 I like it! Usually I write about security here, but Apple's iron-
Cell formatters			000				bound determination to keep Adobe Flash out of any iWhatever device is about to blow up in Apple's face.
O Add cell formatter	December 1st 2021, 11:26:09.642	Mg Siegler	April 4th 2011, 02:00:00.	4	2,011	TechCru nch	A couple weeks ago, after All Facebook the ability to check-in to Facebook events, the social network that the feature would be launching shortly on the iPhone
			000		0.040	7011-1	and their mobile-based touch site. Today, here it is.
Hide time column Customize columns	December 1st 2021, 11:26:09.642	-	r 7th 2013, 02:00:00.	9	2,013	ZUNet	Loougnemy white Nowa Lumia 900 back in May and have designated AT&T as my Windows Phone carrier (I use Verizon for the iPhone and T-Mobile for Android). That was the easy part and for more than a

5. From **Column** (1), select **snippet.** From **Type**, select **NLP** (2).

Implementer Easy Start Tutorial - Siren Platform version 14.7

Visualize / New Visualization (unsaved)							Save Share Re	fresh
Add a filter +								
Data Options	_siren.importTimestamp	author	companies	foundcbref	foundnlpref	pdate	snippet	
Order and visibility The order and visibility of columns can be set in the right panet, click on the @ button to view all the available fields the order (vice on the Markhon to take the	December 1st 2021, 11:26:09.642		company/cisco	true	true	January 10th 2007, 01:00:00. 000		
corresponding column.	December 1st 2021, 11:26:09.642	Sjvn	company/apple	true	true	April 14th 2010, 02:00:00. 000	You think things are bad now between Apple and Adobe? Just wait until the lawsuit. 21 like it Usually I write about security here, but Apple's iron bound determination to keep Adobe Flash out of any iWhatever device is about to blow up in Apple's face.	
Cell formatter 1	December 1st 2021, 11:26:09.642	Mg Siegler	company/facebook,c ompany/the-feature	false	true	April 4th 2011, 02:00:00. 000	A couple weeks ago, after All Facebook the ability to check-in to Facebook events, the social network that the feature would be launching shortly on the iPhone and their mobile-based touch site. Today, here it is.	
Type	December 1st 2021, 11:26:09.642		company/htc,compa ny/nokia,company/ve rizon,company/wind ows,company/windo ws-phone	true	true	Septembe r 7th 2013, 02:00:00. 000	I bought my white Nokia Lumia 900 back in May and have designated AT&T as my Windows Phone carrier (use Verizon for the iPhone and T-Mobile for Android), That was the easy part and for more than a week I have been using the HTC Windows Phone 8X (HTC BX) and Nokia Lumia 920 as I try to figure out which of the two was zoine to be my cersonal choice.	
	December 1st 2021,	Scott	company/chicago-	false	false	July 20th	Chicago Tribune just relaunched its website	

6. From Annotations field, select siren.nlp.instances.snippet, which contains

the annotated offsets needed for formatting the text.

7. Click the **Apply changes** button.

Visualize / New Visualization								Save Share Refre	esh
Add a filter +						_			
Data Options Cell formatter 1 Column shippet Type NI P			phone		02:00:00. 000		Microsoft CEO Steve Balimet played Sante early, promising that all Microsoft employees would get Surface RT tablet/PCs and Windows Phone 8 devices. As reported by Geekwire, the word was all full-time direct, employees (a.k.a.	entity/organization Overlapping annotations	
Anotations field siren.nip.instances.nippet Anotations nested tags field id I Available annotation types entity/phonenumber entity/phonenumber	~	December 1st 2021, - 11:26:09:642	company/apple true	true	Septemb 9 2,0 er 9th 2013, 202:00:00, 000	13 ZDNet	[UPDATE: Thanks to [cd toot for noticing that [cd toot for acknowledged this problem, posted instructions on how to fix it and released and updated version of Thures 8 which can be downloaded as normal. Annoyingly, the update isn't being pushed to all Thures 8 users and if you are affected by	entity/person entity/organization	
Condensed offsets Add cell formatter		December 1st 2021, - 11:26:09.642	company/consultant, true company/federal- communications- commission,compan v/zoogle	true	Septemb 9 2.0 er 7th 2013, 02:00:00. 000	113 ZDNet	The UK's data-protection watchdog has reopened its investigation into Google's unsolicited Street View data	entity/organization	

- 8. Click **Save...**, name the visualization Record Table NLP, and click **Save to dashboard**.
- 9. In the upper-right corner, click **Save > Save dashboard.**

Connecting extracted entities in the data model

Define the relationship between the **articles** index, entities, and the rest of the indices in our data model. To create relations manually, complete the following steps:

1. Go to the **Data model** app.

- 2. Click the **Relations** tab and click **Add relation**.
- 3. Under Source Entity **articles**, from **Field**, select **companies**.
- 4. From Target Entity, select companies and from Field select id.
- 5. Add the labels **mentions** and **mentioned in**.
- 6. Click **Save**.

Example usage of NLP Data in Siren Investigate

Data annotated by using the NLP plugin can be very useful for finding interesting relationships between fields in an index and the annotated text. In this example, you try to find the mentions of companies as a part of the articles snippet.

In the **Data model** app, click the **Data** tab. In the list of **Available fields**, you can see that a new set of fields starting with **siren.nlp** have been created by the NLP plugin during the import. These fields contain annotations for fields processed by Siren NLP (in this case the **snippet** field), such as organization, person, and location.

From the Data model page for the articles index, complete the following steps:

1. On the Relations tab, create a relation between Articles

siren.nlp.matches.entity/organization.keyword and Companies label.

	Fields (195)	Data (646902)	Relations (2)	Scripted fields (0)	Options	Revisions	[™] Dat	ta model graph				
		Source Entit	У		Lab	els			Target Entity			
	ar	ticles		\rightarrow	mentions	5	•		companies	•		
F	ield co	mpanies.keyword	ŀ	- ←	mentione	ed in	•	Field	id	•	×	ļ
	ar	ticles		\rightarrow	mentions	5	•		companies	•		1
F	ield si	ren.nlp.matches.e	ntity/organiz	. ←	mentione	ed in	-	Field	label	•	×	

2. Given that we don't have an index for "Persons", create an Entity Identifier

(EID) and add a relation from it **siren.nlp.matches.entity/person.keyword.**

0	Data	Relations	i ≫ Data model graph						
		S	ource Entity		Labels		Target Entity		
		Persons		\rightarrow	mentioned in	•	articles	•	1

The data model graph should look like this:



3. Go to the **Companies** dashboard and search for Apple ①.

Note: You may need to remove filters from previous exercises.

- To filter the dashboard for the company with the label name Apple, in the label column of the Search Results visualization, click the Filter for value button next to an instance of Apple.
- 5. Go to the **Graph Browser** and drag the **Companies** dashboard there.
- 6. Click the Apple node, right-click, and select **Expand by relation**.
- 7. Select **mentioned in (26)** *articles* and click **Expand**. All the articles that mention this company are displayed.
- 8. Select all nodes on the graph, right-click, and select **Expand by relation**.
- Select mentions (1) Persons and mentions (26) companies and click
 Expand.

You now have an interesting graph that shows articles mentioning Apple and also mentions some **other companies.** The graph also shows mentions of **persons** who are named in articles that talk about Apple.



Creating a Topic Explorer visualization

You can use the Topic Clustering visualization to better explore textual content, for example, understanding what topics are discussed in the articles.

- 1. Open the Articles dashboard.
- 2. From the options menu, click Edit > Add > Add a new visualization
- 3. Select Topic Clustering and select articles.
- In Input Field, select Snippet. The snippet field contains the actual text of the articles.
- 5. Click the **Apply changes** (play) button.
- 6. Click **Save...**, name the visualization Topic Explorer and click **Save**

visualization.

Visua	alize / Ne	w Visualiza	ition (unsave	ed)	
Add a	filter 🕇				
Data	Options	Stop Words	Appearance		×
Input F	ield				
Sni	ippet				~
Root To	pics Extracti	on Algorithm	0		
Clu	istering on	significant te	erms		~
Sub-To	pics Extractio	on Algorithm	3		
Pla	in Significa	ant Terms			~
Chart T	ype				
Po	lygonal Tre	e Map			~



Deleting data and changing a table schema

Deleting data in Siren Platform is intentionally difficult due to the damage that it could cause in environments where the data is in very large streams, such as in large Elasticsearch installations.

The following procedure describes how to delete data. You do not need to complete this for this tutorial.

- 1. Go to the **Dev tools** app.
- In the **Console**, in the editor pane, type delete and the name of the index, for example, siren-.
- 3. Click the play button ①. You see the acknowledgment in the response pane.





If the data structure has changed, for example, if there are more fields or fields of a different type, you must click refresh to refresh the 'Index Pattern Search' field list ①.



Legal notices

Siren Platform™ is a trademark of Sindice Ltd. trading as Siren, with offices in 15 Market St, Galway, H91 TCX3, Ireland.

Elasticsearch™ is a trademark of Elasticsearch B.V., registered in the U.S.A. and in other countries.