Parse & Validate Unstructured Contact Data

OPENPRISE

Cook Book Series

Recipe Overview

This is a recipe to parse name, company, and address data from unstructured source text, leveraging the Google Places API for address validation and parsing

- Set up Google Places API and get an API key
- Create a Google Places API Data Source
- Create a Contact Information Parsing rule to parse unstructured data

You will need the following:

- Have a Google email account
- A data source with a set of fields that contain one or more of the following: Contact Names, Company Names, and Address

tips

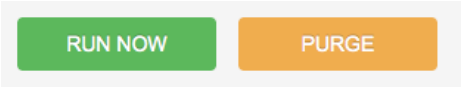
- Add a rule by clicking on an existing rule 

and +.

- Put new data into a new data attribute so you can easily compare before vs. after and confirm the rule is doing what it is supposed to do.

- Can't see the open reference data? Check the setting in your Data Catalog:

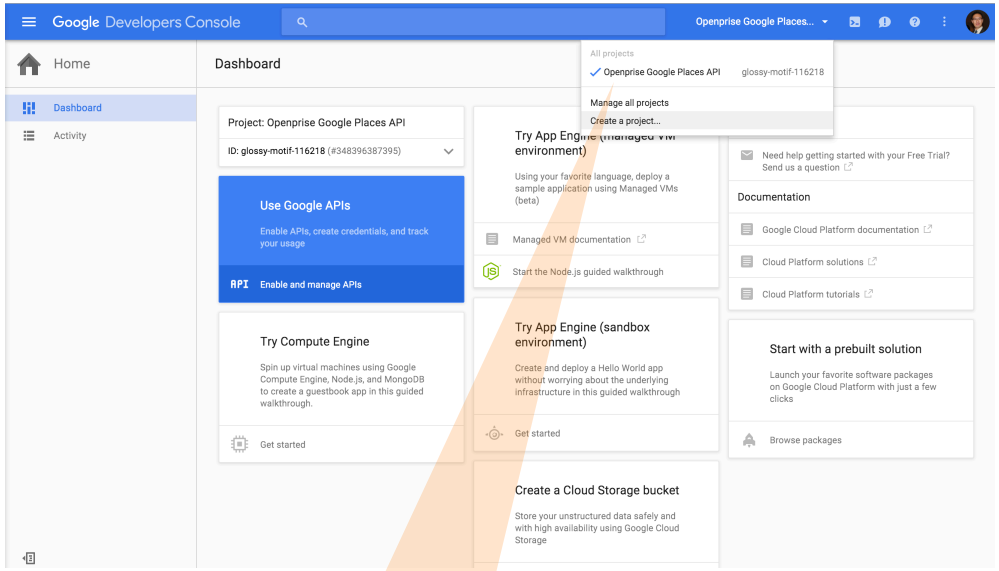
- You can run your Pipeline anytime by pressing the RUN NOW button. To remove all previously processed data within the pipeline, press PURGE first.



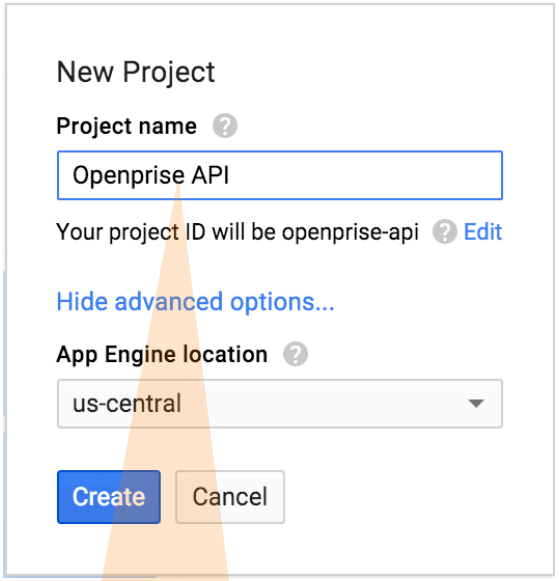
OPENPRISE

Step 1 : Create Developers Console Project

Go to https://console.developers.google.com and login with your existing Google email account if prompted to authenticate.



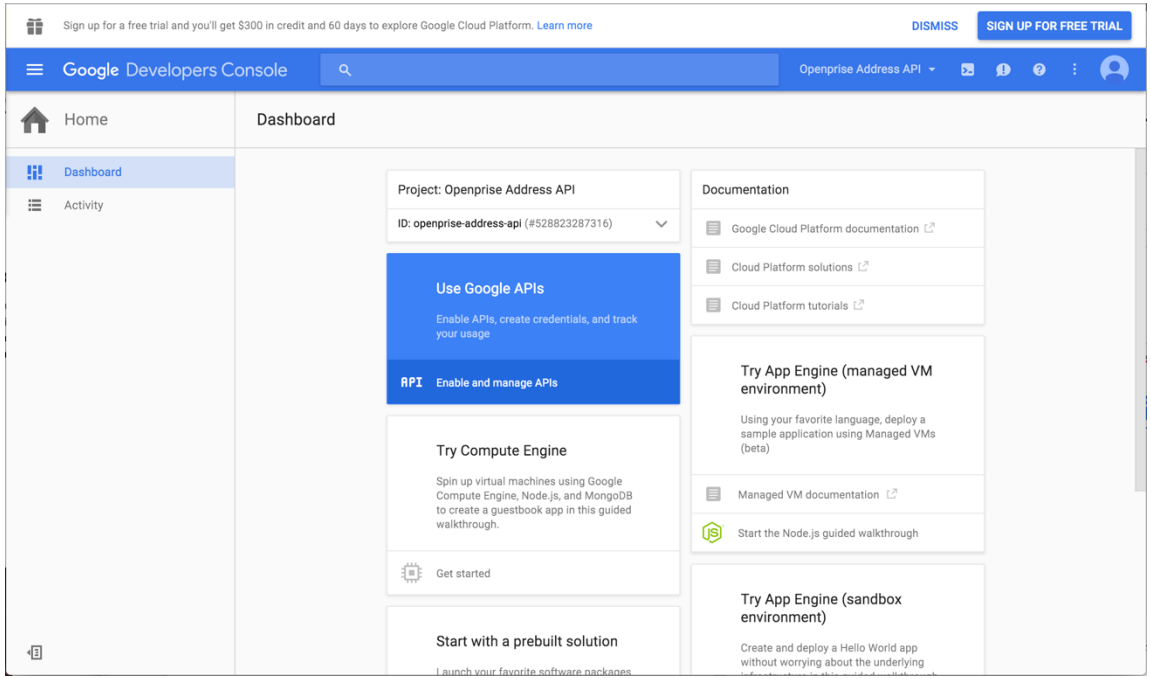
Use the Project dropdown to select “Create a project…” to create a new project for Openprise use.

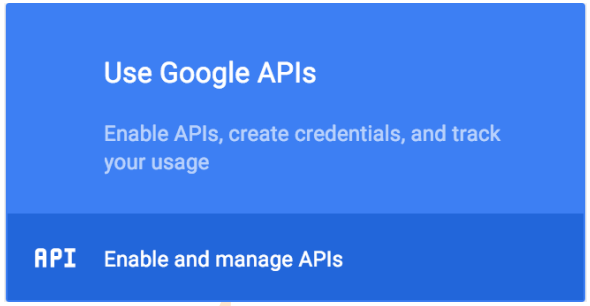


Enter a name for the “Project name” field. For example: Openprise API

OPENPRISE

Step 2 : Enable APIs

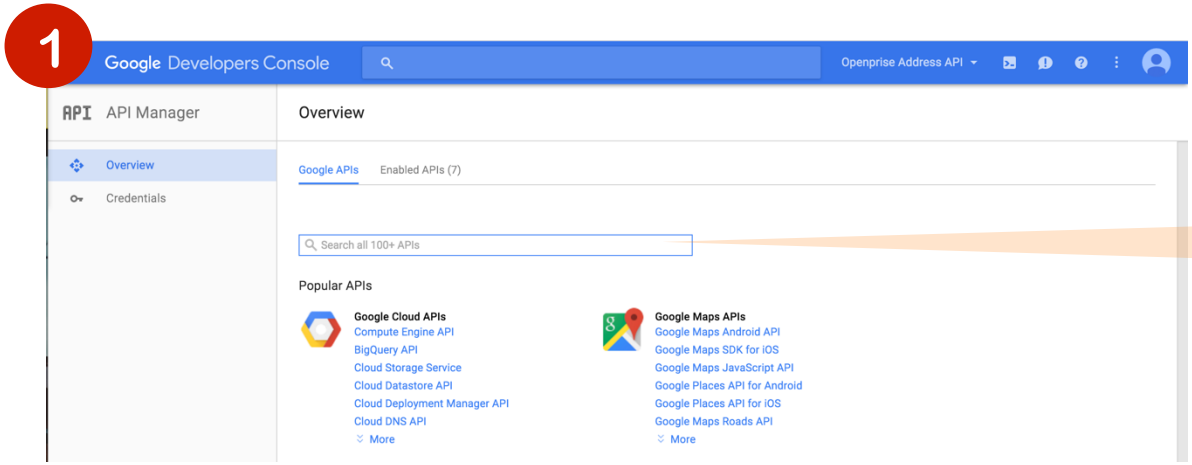




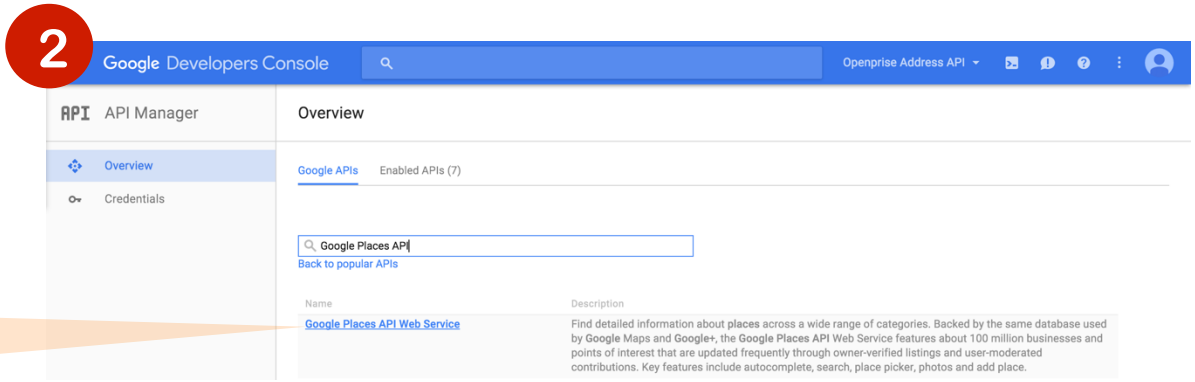
Click on “Enable and manage APIs” to enable Google Places API for this project.

OPENPRISE

Step 3 : Find Google Places API to Enable



Type “Google Places API” in the search field to find the API quickly.

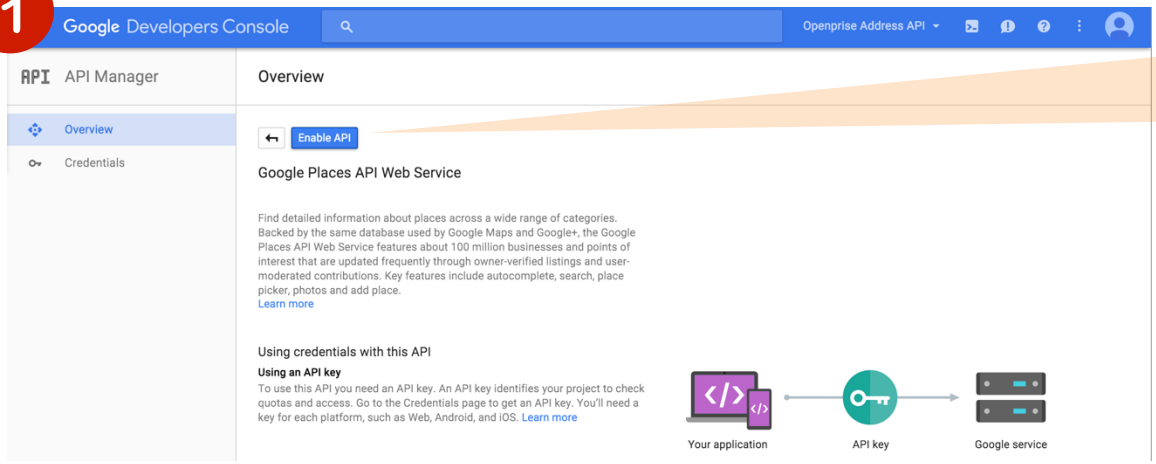


Click on “Google Places API Web Service” search result to enable this particular API for this project.

OPENPRISE

Step 4 : Enable Google Places API

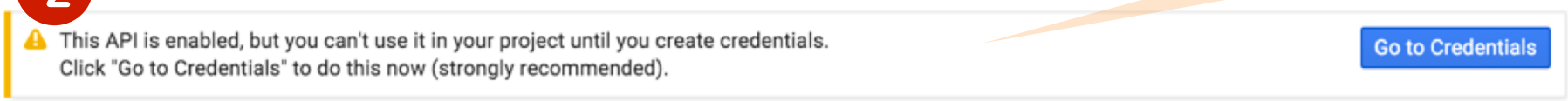
1



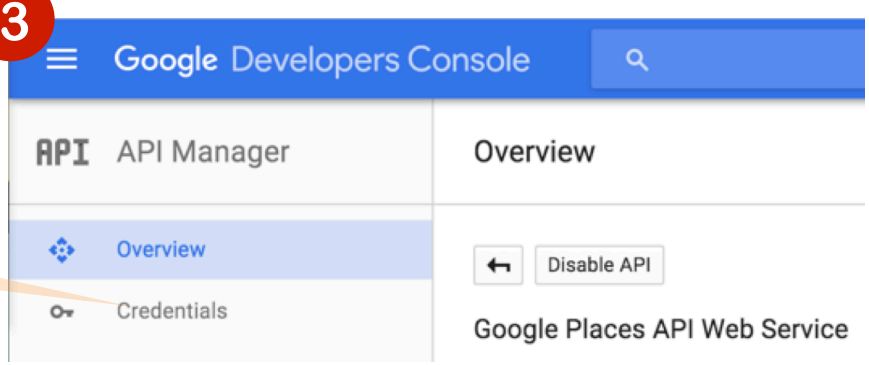
Click on “Enable API” to enable the Google Places API Web Service for this project. Once enabled, you will be prompted to create a credential prior to using this API.

Ignore this message. Do not use the “Go to Credentials” button in this warning message.

2



3



Click on “Credentials” in the left-hand navigation panel instead to create credentials for use with this project.

OPENPRISE

Step 5 : Create API Key Credential

1

APIs
Credentials

You need credentials to access APIs. Enable the APIs you plan to use and then create the credentials they require. Depending on the API, you need an API key, a service account, or an OAuth 2.0 client ID. Refer to the API documentation for details.

New credentials

- API key
- OAuth client ID
- Service account key
- Help me choose

Use the “New credentials” button to create a new credential and pick “API key” as the credential to create

2

Create a new key

You need an API key to call certain Google APIs. The API key identifies your project. Also, it is used to enforce quotas and handle billing, so keep it safe.

Server key | Browser key | Android key | iOS key

Create a “Server key” for use with this project. Openprise requires a server key to connect to your Google Developer’s account.

3

eate server API key

This key should be kept secret on your server

Every API request is generated by software running on a machine that you control. Per-user limits will be enforced using the address found in each request's `userIp` parameter, if specified. If the `userIp` parameter is missing, your machine's IP address will be used instead. Learn more

Name

Server key 1

Accept requests from these server IP addresses (Optional)

Examples: 192.168.0.1, 172.16.0.0/12, 2001:db8::1 or 2001:db8::/64

IP address

Create | Cancel

Enter any name for the name of the API key.

4

API key

Here is your API key

AIzaSyA[redacted]iTjhDc

OK

Your API key is now created. Copy the API key from this screen to use during Data Source creation.

OPENPRISE

Step 6 : Create Billing Account (Optional)

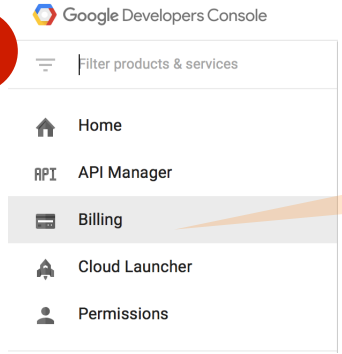
Google Places API Web Service enforces a default limit of 1,000 requests per 24 hour period. It can be increased to 150,000 requests per 24 hour period if a user verifies identity through enabling a billing account. This does not mean you will be automatically charged for usage. You can set daily quota limit to control spent.

For more information on Google Places API Web Service usage limits:
https://developers.google.com/places/web-service/usage?hl=en_US

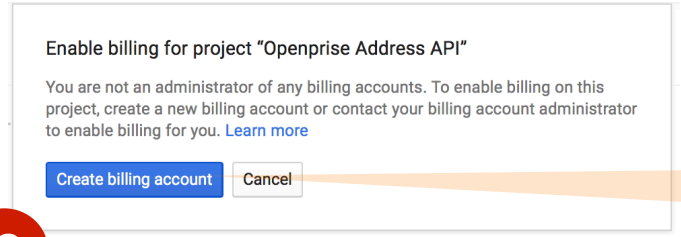
1

Click on the “burger menu” icon in the top left corner of the Google Developers Console

2

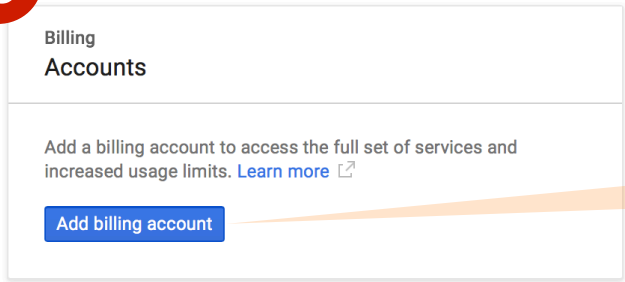


Click on “Billing” to access the billing page



If this is the first project you created in the Developers Console, you will be prompted with this and you can click on “create billing account directly.

3



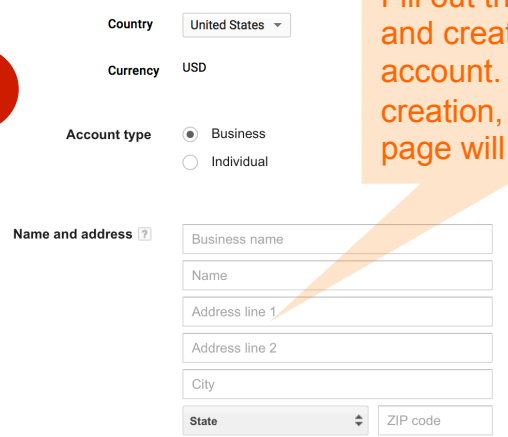
If you have other projects or billing accounts, you can just “Add billing account” to create a new one.

OPENPRISE

Step 6 : Create Billing Account (Continued)

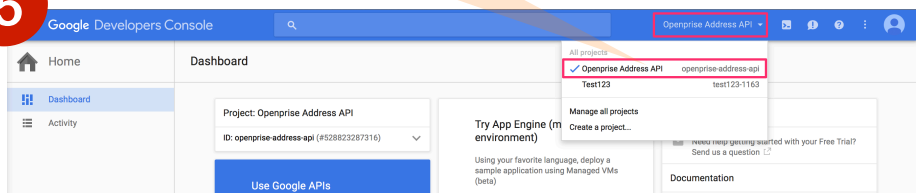
Google is offering a $300 credit as a free trial currently (as of December 2015) and adding a billing account enables you to utilize this offer.

4



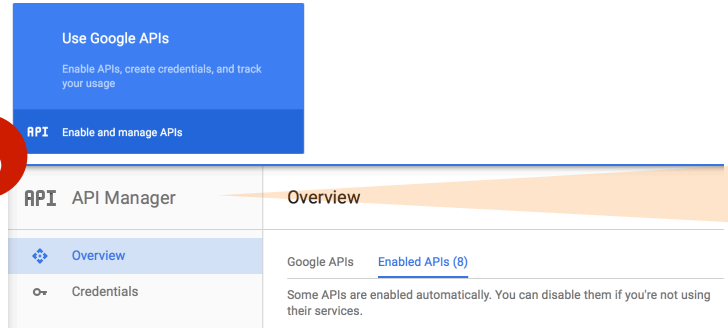
Fill out the billing form and create the billing account. After creation, the Home page will be shown.

5



Make sure you are selecting the project that we have created for Openprise earlier

6



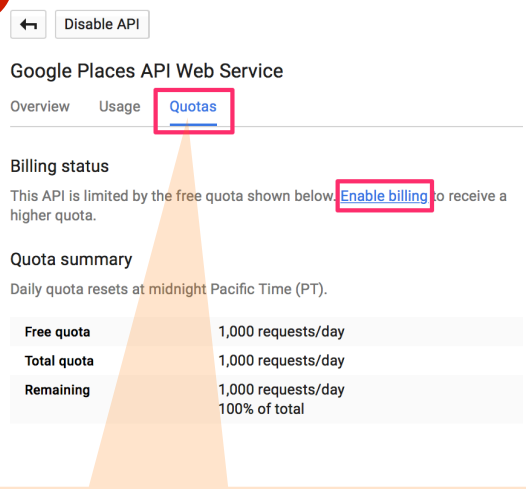
Go back to the enabled Google Places API Web Service page via “Enable and Manage APIs” then click on “Enabled APIs” and click on “Google Places API Web Service”

OPENPRISE

Step 6 : Create Billing Account (Continued)

Google is offering a $300 credit as a free trial currently (as of December 2015) and adding a billing account enables you to utilize this offer.

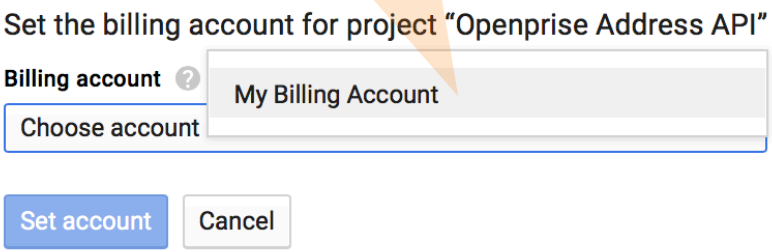
7



Now click on “Quota” to see the current quota. And then click on “Enable billing” to link the created Billing Account to this project

Now from the “Choose account” dropdown, pick the billing account you created to link to this project.

8

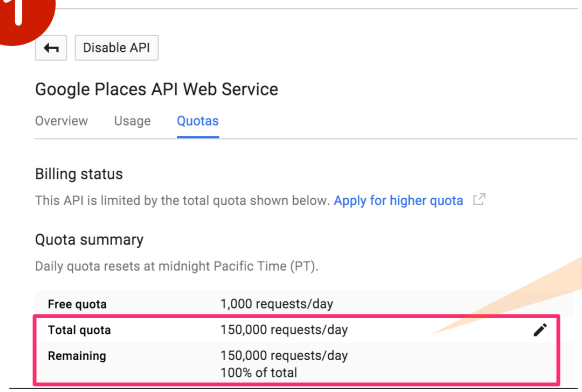


OPENPRISE

Step 7 : Set Quota Limit (Optional)

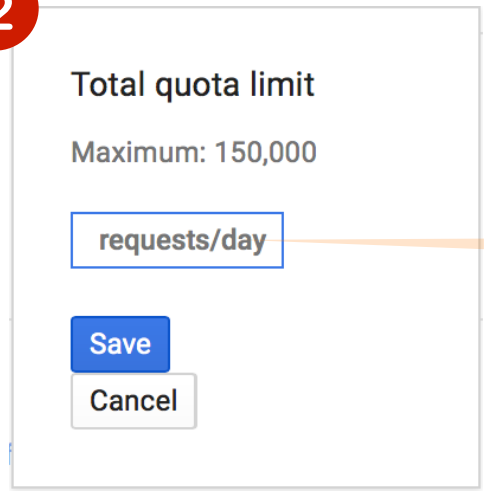
Openprise may use more than 1 API call for each processed record. If there are overcharge concerns, Google Developers Console allows users to set quota limit on the APIs.

1



On the Google Places API Web Service page, the Quota section shows the existing quota limits and usage. Next to "Total quota", you can edit the Total Quota for the API

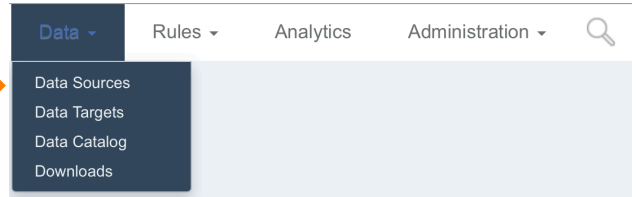
2



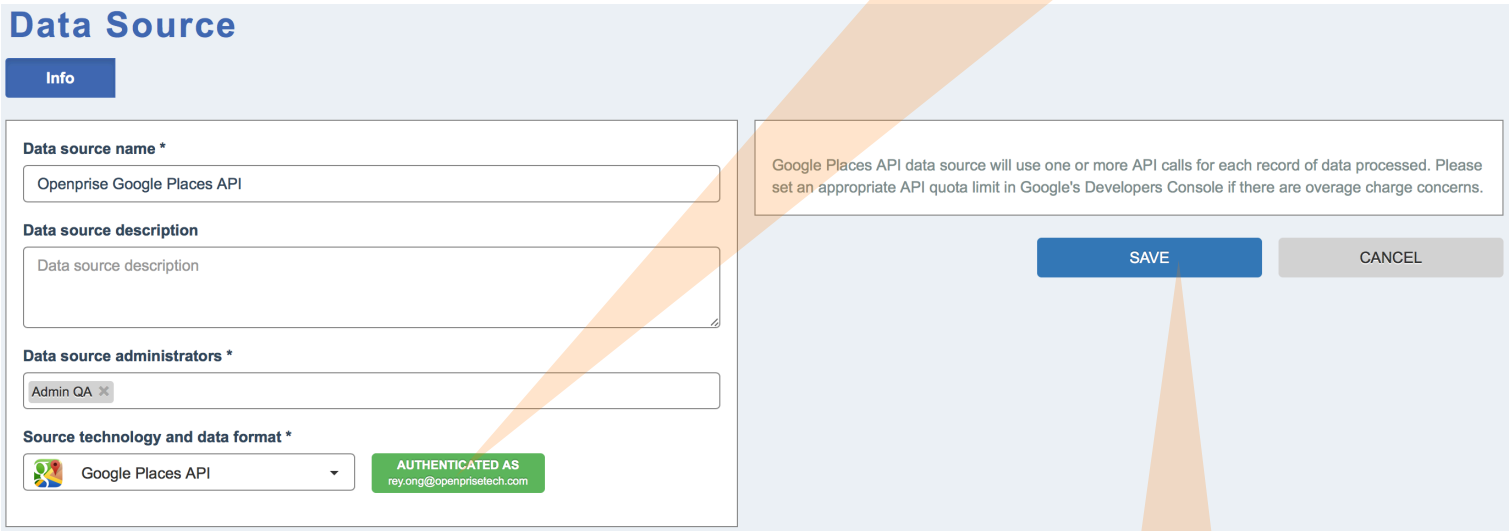
Set an appropriate limit for your usage

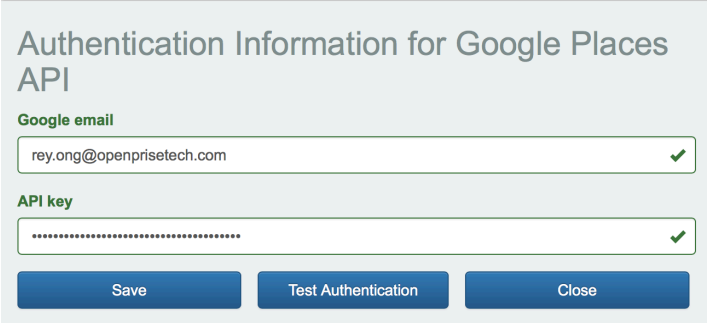
OPENPRISE

Step 8 : Google Places API Data Source



Make sure you authenticate using the email account and API key you created in Step 5 of this cookbook. You can use “Test Authentication” to verify that the authentication works before “Save”.





Save the data source

Need more help creating a Data Source? Check out the tutorial videos on Openprise website’s Resource page: http://openprisetech.com/resources

OPENPRISE

Step 9 : Contact Information Parsing Rule

IF this happens

Input Data Sources

Contact Information Testing

SELECT DATA

All new data

Use the Contact Information Parsing rule template to parse out contact names, company names, and addresses from a set of fields

Contact Information Parsing

Add all attributes that may contain contact information (people names, company name and/or address). Please order the attributes in sequential order for processing to increase parsing and validation accuracy.

ADD ANOTHER FIELD

Attribute1	1ST LINE OF NAME & ADDRESS
Attribute2	2ND LINE OF NAME & ADDRESS
Attribute3	3RD LINE OF NAME & ADDRESS
Attribute4	4TH LINE OF NAME & ADDRESS
Attribute5	5TH LINE OF NAME & ADDRESS

Select as many fields that contain Contact Information. In this particular sample, there are 5 lines that may contain contact names, company names and address.

- [x] **Parse out contact names** Show output attribute names
- [x] **Parse out company name** Show output attribute names
- [x] **Parse out address** Show output attribute names

Google Places data source Google Places API EK1217

No. of address to retrieve 2

(Note:- This controls the number of full address fields that we return from Google Places API)

No. of address to parse 2

(Note:- This controls the number of addresses that will be parsed into address components like street number, street, city, state, country, etc...)

Choose what information to parse out. Google Places Data Source that was created in Step 8.

Show advanced configuration

Advanced configurations

For better parsing accuracy, street suffix keywords and company name keywords are required.

Company keywords are stored in	Reference - Company Name Clean Up
Values are stored in	Company Name Keywords
Street suffix alias/keywords are stored in	Reference - Street Suffix Abbreviations
Values are stored in	Alias

Show less

Use the advanced configurations if you want to fine tune the results

OPENPRISE

Person Names Parsed From A Set of Fields

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
	1ST LINE OF NAME & ADDRESS ▲ ABC	2ND LINE OF NAME & ADDRESS ABC	3RD LINE OF NAME & ADDRESS ABC	4TH LINE OF NAME & ADDRESS ABC	5TH LINE OF NAME & ADDRESS ABC	6TH LINE OF NAME & ADDRESS ABC	7TH LINE OF NAME & ADDRESS ABC	OP_Salutation1 ABC	OP_First_Name1 ABC	OP_Middle_Name1 ABC	OP_Last_Name1 ABC	OP_Suffix1 ABC	OP_Full_Name1 ABC	OP_Salutation2 ABC	OP_First_Name2 ABC	OP_Middle_Name2 ABC	OP_Last_Name2 ABC	OP_Suffix2 ABC	OP_Full_Name2 ABC
1																			
2																			
3																			
4																			
5																			
6																			
7																			
8																			
9																			
10																			

Company / Organization Names Parsed From A Set of Fields

	A	B	C	D	E	F	G	H
	1ST LINE OF NAME & ADDRESS ▲ ABC	2ND LINE OF NAME & ADDRESS ABC	3RD LINE OF NAME & ADDRESS ABC	4TH LINE OF NAME & ADDRESS ABC	5TH LINE OF NAME & ADDRESS ABC	6TH LINE OF NAME & ADDRESS ABC	7TH LINE OF NAME & ADDRESS ABC	OP_Company ABC
1								
2								
3								
4								
5								
6								
7								
8								
9								
10								

OPENPRISE

Addresses Parsed From A Set of Fields

	A	B	C	D	E	F	G	H	I	J	K	L	M
	1ST LINE OF NAME & ADDRESS ▲ ABC	2ND LINE OF NAME & ADDRESS ABC	3RD LINE OF NAME & ADDRESS ABC	4TH LINE OF NAME & ADDRESS ABC	5TH LINE OF NAME & ADDRESS ABC	6TH LINE OF NAME & ADDRESS ABC	7TH LINE OF NAME & ADDRESS ABC	OP_Parsed_Address ABC	OP_ADD1_Street_Number ABC	OP_ADD1_Street_Address ABC	OP_ADD1_State ABC	OP_ADD2_Country ABC	OP_ADD1_Country ABC
1	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]		[illegible]	[illegible]	[illegible]	[illegible]		[illegible]
2	[illegible]	[illegible]	[illegible]	[illegible]				[illegible]	[illegible]	[illegible]	[illegible]		[illegible]
3	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]		[illegible]	[illegible]	[illegible]	[illegible]		[illegible]
4	[illegible]	[illegible]	[illegible]					[illegible]	[illegible]	[illegible]	[illegible]		[illegible]
5	[illegible]	[illegible]	[illegible]	[illegible]				[illegible]	[illegible]	[illegible]	[illegible]		[illegible]
6	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]			[illegible]	[illegible]	[illegible]	[illegible]		[illegible]
7	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]			[illegible]	[illegible]	[illegible]	[illegible]		[illegible]
8	[illegible]	[illegible]	[illegible]					[illegible]	[illegible]	[illegible]	[illegible]		[illegible]
9	[illegible]	[illegible]	[illegible]	[illegible]				[illegible]	[illegible]	[illegible]	[illegible]		[illegible]

OPENPRISE

Recipe Review

Recommendations

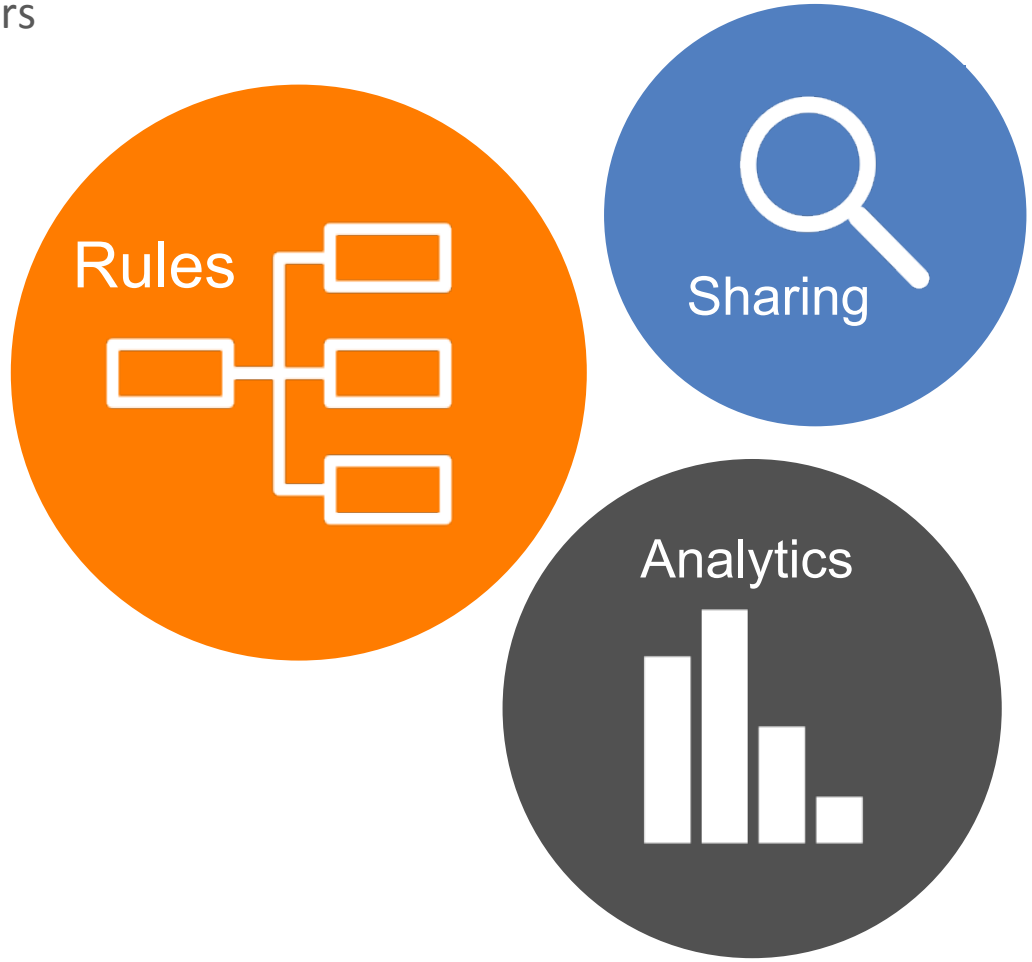
- Remember to add Billing Account to increase the daily quota and take advantage of Google's free trial with $300.00 credit.
- Fine tune the Contact Information Parsing rule using your own Street Suffix keywords and Organization keywords. Download the Open Data then upload the modified versions as your own.

Want to do more? Try the following on your own:

- Clean and normalize the data after parsing, for example: normalize state and country names, clean up company names
- Infer missing data, for example: infer missing city and state data from postal code data

OPENPRISE

Data Automation For Business Users



info@openprisetech.com

Twitter: @openprisetech

www.openprisetech.com