

Following the Money: A Beginner's Guide to

August 24, 2023

Finance OpenCorporates

Open source research isn't only about analysing social media or satellite imagery. Another important area involves investigating company structures and relationships.

However, official company registries can be unwieldy or difficult to navigate.

This is where OpenCorporates – a free repository of company registries aggregated from primary public sources, published in order to promote corporate transparency – can be helpful. OpenCorporates is clear about the provenance of their data, so you know where they collected each record, and when. This is critical for companies and finance research and investigations.

While there are other databases like it, OpenCorporates is the only one that covers so many jurisdictions (145 as at 16 June 2023). To see other sites that offer access to corporate registries, go to the companies and finance tab of the [Bellingcat Online Investigation Toolkit](#).

Still, even user-friendly websites like OpenCorporates can be a daunting prospect – conducting searches one by one takes a great deal of time, and it can involve a lot of copying and pasting of search results. But the Open Corporates API can save a lot of time.

What is an API and Why Use One?

An API is an Application Programming Interface. Broadly speaking, it allows users to obtain data from a database without having to know about the structure or languages used to manage that database. Users can build new and interesting tools with API access or create different ways to look at the data they contain, as many previously did with the Twitter API, for example.

Put simply, if you're researching data at scale, access to the API of an organisation that stores a lot of data can be a powerful resource. It can allow researchers to explore and compare data in ways that might not otherwise be possible.

Imagine, for example, that you're looking at a company and all the individuals and other companies related to it within OpenCorporates. If you did individual searches for each company and director, it would take a lot of note-taking and organising. Access to an API can allow you to pull all the data at once and create new methods of searching to more easily identify relationships.

Luckily, OpenCorporates provides free API access if you're undertaking a public benefit project. We'll detail how to apply for access later on in this guide.

Some investigative journalists say this method has helped them find leads earlier. David Szakonyi, co-founder of the Anti-Corruption Data Collective, reflects that "by accessing OpenCorporates' API, we achieved in less than a day what would have taken two people between four and six months to do". OpenCorporates helped the ICIJ to connect companies and directors for over 240,000 companies as part of the Panama Papers investigation, before releasing it all as open data.

What follows is a guide on how to get the best from Open Corporates. If you have coding or tool-building experience, you will be able to do more. But even those who don't can learn how to use it and gather all manner of potentially valuable and revealing information. What'smore, once you understand how to explore the OpenCorporates API, you can begin to transfer that knowledge to other large datasets that provide API access such as the OpenSanctions database.

Much of what is contained within this guide can also be explored in the below video by Rebecca Lee of OpenCorporates. But this guide aims to build upon this very useful resource.

#CIJSUMMER Workshops: OpenCorporates API for Beginners



Editor's note: Some of the images contained in this guide have been blurred as they include the details of real companies that are featured on OpenCorporates.

Set up your Free OpenCorporates Account

Go to the OpenCorporates (OC) [service desk](#) to apply for access as a public benefit project. Academics, NGOs, registered journalists, media organisations and registered nonprofits can apply for free access under an open public licence. If you're an independent journalist, you'll need to use your own name and email address and provide a list of media sites that have referred to your reporting. OpenCorporates may ask for a copy of a current press pass (or equivalent).

Note that their licence requires you to attribute the use of their database in your reporting or published investigation outputs — see their [terms of use](#).

Choose your Method

Based on the needs of your investigation, it is possible to obtain:

- Data for one company
- Data for one company name across all jurisdictions
- Data for multiple companies with different names

If you're not sure yet, start with the 'Data for one company' section as it's the easiest, then work your way up from there.

OpenCorporates has a [knowledge base](#), which is a good place to find things like the API documentation and [data dictionaries](#) (which cover things like what is the 'normalised_name' field – you need to understand this when you are drawing data to use in an investigation).

Data for One Company (Browser & Excel)

Use this method when you have no programming experience. It's useful when you want to pull the search results for one company into Excel. But it does have the limitation that Excel is made for viewing data all in one table.

This method involves: Using your browser to query the API; downloading an XML file; and viewing the data in Excel by opening it as an XML table.

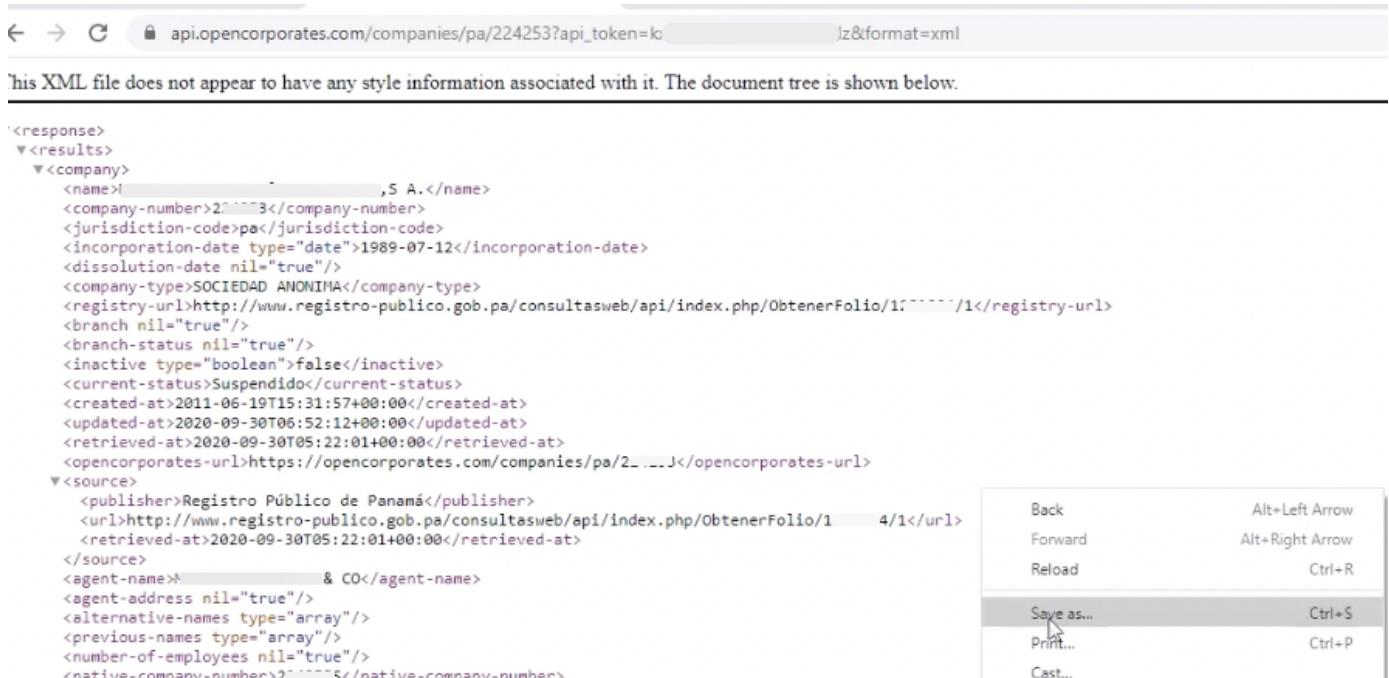
Query the API With your Browser

Search for the company that you want to view on opencorporates.com and select it (i.e. click on the company name in the search results to view its page).

Go to the browser's address bar and adjust the URL:

- Add api at the front of the URL.
- Add ?api_token= at the end of the URL.
- Add the API key you received via email to the end of the URL.
- Add &format=xml to the end of that.
- You should end up with a URL something like this:
http://api.opencorporates.com/companies/pa/company_number?api_token=XXXX&format=xml
(Replace 'company_number' with the company number that you want to query.)

Hit enter, and the data you have queried via the API will appear in your browser. Right-click and save as page source:



A screenshot of a web browser displaying the XML page source of a company. The URL in the address bar is `api.opencorporates.com/companies/pa/224253?api_token=kJz&format=xml`. The page content shows an XML document with company details like name, number, incorporation date, and status. A context menu is open on the right side of the screen, with the 'Save as...' option highlighted.

```

<response>
  <results>
    <><company>
      <name>S. A. </name>
      <company-number>211113</company-number>
      <jurisdiction-code>pe</jurisdiction-code>
      <incorporation-date type="date">1989-07-12</incorporation-date>
      <dissolution-date nil="true"/>
      <company-type>SOCIEDAD ANONIMA</company-type>
      <registry-url>http://www.registro-publico.gob.pa/consultasweb/api/index.php/ObtenerFolio/171111111</registry-url>
      <branch nil="true"/>
      <branch-status nil="true"/>
      <inactive type="boolean">false</inactive>
      <current-status>Suspendido</current-status>
      <created-at>2011-06-19T15:31:57+00:00</created-at>
      <updated-at>2020-09-30T06:52:12+00:00</updated-at>
      <retrieved-at>2020-09-30T05:22:01+00:00</retrieved-at>
      <opencorporates-url>https://opencorporates.com/companies/pa/211113</opencorporates-url>
    </company>
    <><source>
      <publisher>Registro Público de Panamá</publisher>
      <url>http://www.registro-publico.gob.pa/consultasweb/api/index.php/ObtenerFolio/1</url>
      <retrieved-at>2020-09-30T05:22:01+00:00</retrieved-at>
    </source>
    <agent-name>N & CO</agent-name>
    <agent-address nil="true"/>
    <alternative-names type="array"/>
    <previous-names type="array"/>
    <number-of-employees nil="true"/>
    <native-company-number>211113</native-company-number>
  </results>
</response>

```

Back	Alt+Left Arrow
Forward	Alt+Right Arrow
Reload	Ctrl+R
Save as...	Ctrl+S
Print...	Ctrl+P
Cast...	

A screen grab from OpenCorporates. Company names and numbers have been partially blurred.

View the XML File in Excel

- Open the XML file in Excel, and if presented with an option, choose to open as an XML table.
- You might get two error boxes that pop up — just click OK.

You will need to tidy the table as when it displays the data, Excel will duplicate some of the company information (in columns) as it creates one row per officer. You will also need to remove some columns, like '@type'.

Data for One Company Name Across all Jurisdictions (Browser & Excel)

Use this method when you have no programming experience but are comfortable in Excel. It's useful when you know the company name and want to draw data for all the companies with that name across all the jurisdictions that OpenCorporates covers.

This method involves: Using your browser to query the API; downloading an XML file; viewing the data in Excel by using the XML source task pane.

Query the API with your Browser

Search with the company name on opencorporates.com. You should get a URL that looks something like this:

<https://opencorporates.com/companies?q=company+name&utf8=%E2%9C%93> (Your search terms will appear in place of 'company' and 'name'.)

Go to the browser's address bar and adjust the URL:

- Add api at the front of the URL.
- Add /search after /companies
- Remove &utf8=%E2%9C%93
- Add &api_token= at the end of the URL.
- Add the API key you received via email to the end of the URL.
- Add &format=xml to the end of that.
- You should end up with a URL that looks like this:

https://api.opencorporates.com/companies/search?q=company+name&api_token=XXXX&format=xml

(Note: Replace XXXX with your API token.)

Hit enter and the data you have queried via the API will appear in your browser. Right-click and save as XML.

View the XML File in Excel – Open Using the XML Source Task Pane

Depending on which version of Excel you have, opening the file in Excel will work differently.

It's also possible to upload to Google Sheets — an explanation of how to do so can be found [here](#).

Microsoft 365 on a Mac

Open the file and it will open like any Excel file. Then clean the data to make it more usable (e.g. delete unnecessary columns).

Windows' version of Excel

Open the XML file, and make sure that you choose 'Use the XML source task pane'.

Click 'OK' on the next box that pops up. You should see something like this:

The screenshot shows the 'XML Source' ribbon tab selected in Excel. A dropdown menu shows 'response_Map'. The main pane displays the XML schema structure:

- response
 - results
 - companies
 - type (array)
 - company
 - name (C [REDACTED] LIMITED)
 - company-number (02 [REDACTED] 5)
 - jurisdiction-code (gb)
 - incorporation-date
 - <value> (1994-06-23)
 - type (date)
 - nil (true)

A screen grab from OpenCorporates. Company names and numbers have been partially blurred.

This is a schema based on the XML data that you have opened — it's the structure of the data.

Select Which Fields you Want to View

Select the fields by dragging and dropping them onto the Excel sheet:

The screenshot shows a worksheet with columns A through H. Row 1 contains the word 'name' in cell A1. A blue selection bar is visible under the 'name' cell, indicating it is being dragged. To the right, the 'XML Source' ribbon tab is selected, showing the schema structure. In the schema, the 'name' node is highlighted in blue.

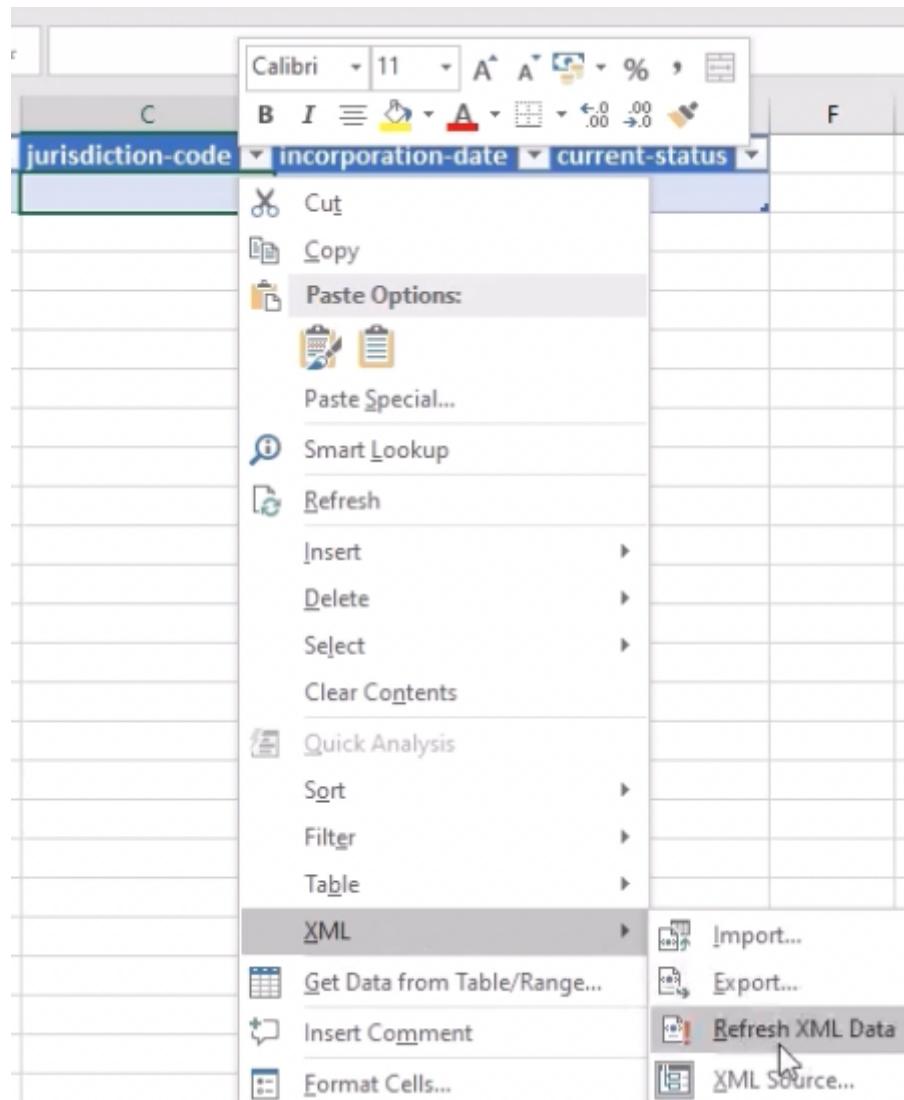
A screen grab from OpenCorporates. Company names and numbers have been partially blurred.

Drag and drop as many fields as you need as columns in row 1. You'll end up with something like this:

The screenshot shows the same worksheet with multiple fields now in row 1: 'name', 'company-number', 'jurisdiction-code', 'incorporation-date', and 'current-status'. Each field has a dropdown arrow next to it, indicating they are dynamic range references.

A screen grab from OpenCorporates.

To populate the sheet with your chosen data, right-click somewhere on row 1, and choose 'Refresh XML data':



A screen grab from OpenCorporates. Company names and numbers have been partially blurred.

That will pull in the data for the fields you want. It will look something like this:

	A	B	C	D	E
1	name	company-number	jurisdiction-code	incorporation-date	current-status
2	LIMITED	0254225	gb	1994-06-23	Dissolved
3	LIMITED	017	hk	1989-08-11	Dissolved
4	& ASSOCIATES,S A.	23	pa	1989-07-12	Suspendido
5	(U.K.) LIMITED	833	bs		Struck Off
6	(BALTICS) LIMITED	014	gb	2010-04-26	Dissolved
7	(CANADA) LIMITED	039	gb	2010-04-23	Dissolved
8	(MOLDOVA) LIMITED	011	gb	2010-04-23	Dissolved
9	, LIMITED	010	gb	2010-04-26	Dissolved
10	(UKRAINE) LIMITED	0114	gb	2010-04-26	Dissolved
11	ECUADOR S.A.	715	pa	2012-01-19	Disuelto
12	SA (SUCCURSALE LUGANO)	1000011	ch		Gelöscht
13	(ASIA) LIMITED	0105	hk	2002-09-16	Dissolved
14	(BAHAMAS) LIMITED	370710	bs		
15	(BRASIL) S/C LTDA	013	br		
16	(H.K.) LIMITED	015	hk	1993-11-25	Dissolved
17	(HONG KONG) INCORPORATED	113	pa	1987-01-22	Vigente
18	(JERSEY) LIMITED	E15	je	1992-07-02	Dissolved
19	(MALTA) LTD	015	mt	2013-05-23	
20	NEW ZEALAND LIMITED	425202	nz	2012-12-15	Removed

A screen grab from OpenCorporates. Company names and numbers have been partially blurred.

Note: You may well end up with less results in your Excel sheet than in your search on opencorporates.com. This is because there is a default page limit in the XML results (see the bottom of the page in your browser):

```
<per-page type="integer">30</per-page>
<total-pages type="integer">2</total-pages>
<total-count type="integer">51</total-count>
</results>
```

A screen grab from OpenCorporates.

To get all the results, you need to add this to the URL that you used to query the API:
`&per_page=100&page=1`

Your URL will look something like this:

https://api.opencorporates.com/companies/searchq=company+name&api_token=XX_XX&format=xml&per_page=100&page=1

Hit enter to refresh the page, and save that page as XML. If your search results on the OpenCorporates website were more than 100, just change the end of the URL to `&page=2` and hit enter to get the next page of data.

Data for Multiple Companies with Different Names (XML & OpenRefine)

Use this method when you have a list of companies for which you want to find matching legal entities across all jurisdictions on OpenCorporates. This method is more involved and requires the use of OpenRefine, a popular open source data wrangling tool that has built-in functions for accessing data through APIs, something that is not possible with Excel or other spreadsheet programs. Naturally, this will involve more learning if OpenRefine is new to you. But it's worth it if you want to draw down company data at scale – this is where using the API gets really powerful!

A note: In this method you are using the OpenCorporates reconciliation API (a.k.a. the legal entity resolution API).

This method involves:

- Getting further approval from OpenCorporates
- Installing OpenRefine
- Reconciling (or matching) your list against the OpenCorporates database
- Improving your reconciliations
- Pulling data into OpenRefine
- Parsing the fields that you want.

Get Approval from OpenCorporates

If you want to access the API via OpenRefine you will need to get on the “approved” list by contacting community@opencorporates.com.

Install and Set Up in OpenRefine

- Go to openrefine.org.
- See OpenRefine’s installation requirements and instructions
- Create a CSV file with the list of companies that you want to query against OpenCorporates. The list of companies can come from research that you have already been doing on OpenCorporates or from any other list or source you may have gathered. It’s better to have more columns in your CSV as this will help OpenRefine to find better matches in the database. In the example used here, the columns are company name, country code, and date active. To create a CSV, put your data in Excel and save as CSV
- Create a project in OpenRefine, click ‘Choose Files’ and upload the CSV
- Click ‘Next’ and you should see something like this:
 -

The screenshot shows the OpenRefine interface with a table of company data. The table has columns for Name, Country code, and Date active. The first few rows of data are:

	Name	Country code	Date active
1.	Private Equity LLC	us_ma	15/05/2006
2.	HOLDING ApS	dk	29/07/2020
3.	; L.P.	us_va	16/03/2015
4.		us_co	15/05/2008
5.	Holdings	nl	15/05/2019
6.	Limited	nz	14/01/2007
7.	GmbH	de	16/05/2011
8.	U.K. Limited	gb	16/05/2011
9.	Inc.	us_de	16/05/2007
10.	Enterprises	us_tx	14/06/2020

A screen grab from OpenCorporates. Company names and numbers have been partially blurred.

- Click ‘Create Project’ (at the top right of the screen). You are now ready to reconcile your list of companies with the OpenCorporates database.

Reconcile (Match) your List against OpenCorporates

- Reconciling is the matching of your list of companies to companies in the OpenCorporates database, so that you pull down the data for the right companies.
- You need to tell OpenRefine which column to use to reconcile against the OpenCorporates database. To do this, click on the arrow next to the ‘Name’ column, scroll down to ‘Reconcile’ and select ‘Start reconciling’:

10 rows

Show as: rows records Show: 5 10 25 50 rows

All	Name	Country code	Date active
Facet	Equity LLC	us_ma	15/05/2006
Text filter	pS	dk	29/07/2020
Edit cells		us_va	16/03/2015
Edit column	tainer	us_co	15/05/2008
Transpose	ngs	nl	15/05/2019
Sort...		nz	14/01/2007
View		de	16/05/2011
Reconcile	K. Limited	gb	16/05/2011
		us_de	16/05/2007
	Start reconciling...		14/06/2020

Facet Reconcile text in this column with items in another data source

Actions

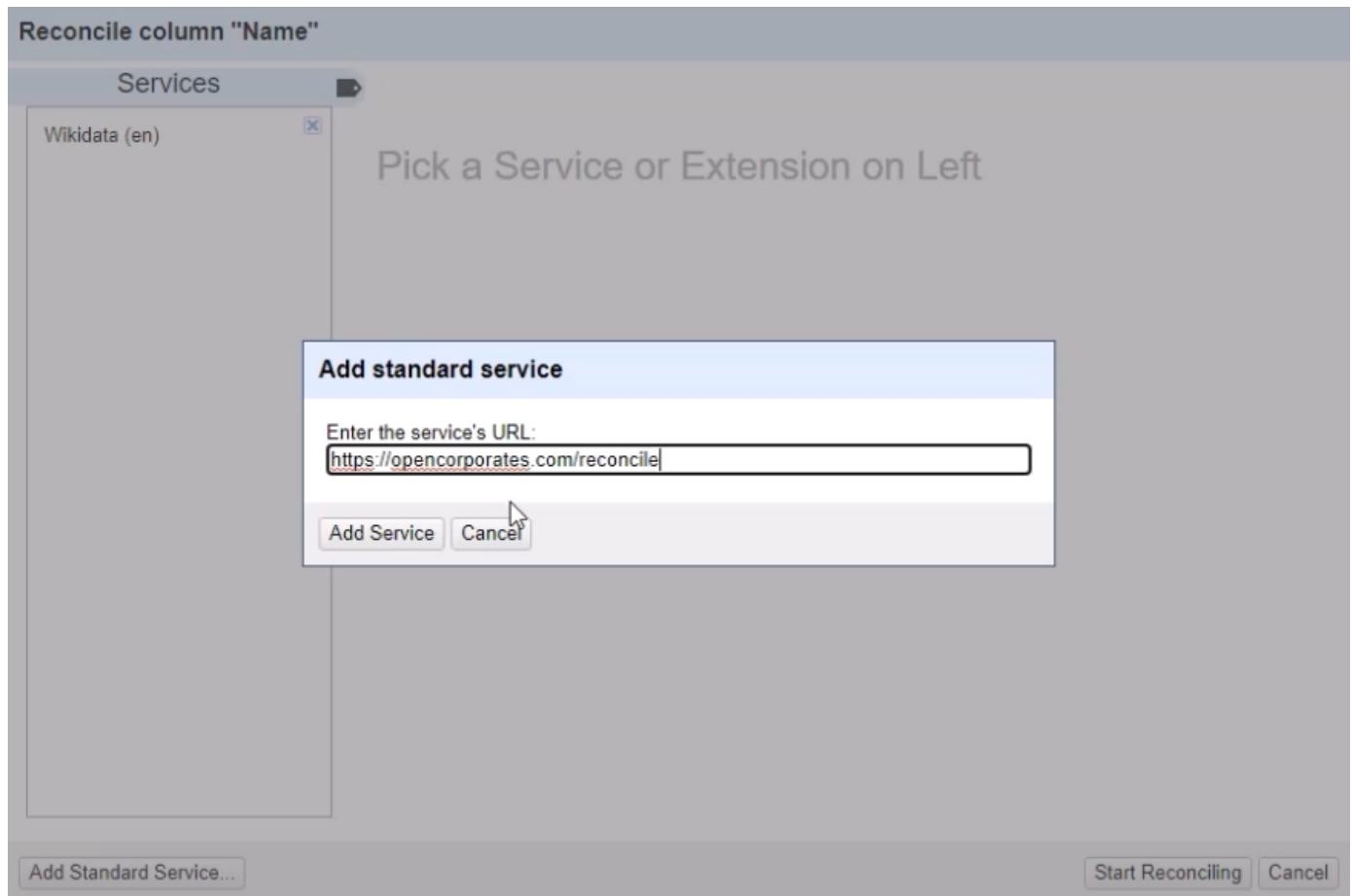
Copy reconciliation data...

Use values as identifiers

Add entity identifiers column

A screen grab from OpenCorporates.

- Click on ‘Add Standard Service’ in the box that pops up in order to add the OpenCorporates service.
- Enter this URL: <https://opencorporates.com/reconcile> and click ‘Add Service’:



A screen grab from OpenCorporates.

(Note: This URL may change. If it doesn't work, check the API documentation on api.opencorporates.com.)

- Click 'Start Reconciling'.
- Once reconciling is finished, the data will look like this in OpenRefine:

10 rows					
Show as:		rows	records	Show:	5 10 25 50 rows
	All	Name	Country code	Date active	
1.		Private Equity LLC PRIVATE EQUITY LLC (50) PRIVATE EQUITY INC (18)	edit us_ma	15/05/2006	
		<input checked="" type="checkbox"/> Create new item			
		Search for match			
2.		HOLDING ApS ApS (35)			
		<input checked="" type="checkbox"/> Create new item			
		Search for match			
3.		L.P. L.P. (31)			
		<input checked="" type="checkbox"/> Create new item			
		Search for match			
4.		Ball Metal Beverage Container E CONTAINER CORP. (38) CONTAINER CORP. (26) CONTAINER Corp. (26)			
		<input checked="" type="checkbox"/> Create new item			
		Search for match			

A screen grab from OpenCorporates. Company names and numbers have been partially blurred.

- The companies in blue (although partially blurred in the image above) are the ones that OpenRefine has reconciled (matched) with in the OpenCorporates database.
 - The reconciliation score is in brackets in grey — that is a score for how close the match is. (It's out of 100.) In this case the match was on company name, which was the column chosen above to reconcile
 -
 - If you hover over a matched company name, the data from OpenCorporates will pop up — this will help you to assess if it's a good match. It is important to assess this yourself as OpenCorporates does not guarantee that the results it pulls up are a match.
 - You can manually select the right match for each company in your list (your CSV), by clicking on the tick next to the company that you think is the best match.
 - Alternatively, you can match the companies in your list to all the highest-scoring matches from OpenCorporates. To do this, click on the arrow next to 'Name' (or

whatever the column header was in your data), go down to ‘Reconcile’, then ‘Actions’, then select ‘Match each cell to its best candidate’:

10 rows

Show as: rows records Show: 5 10 25 50 rows

All	Name	Country code	
1.	Facet Text filter	RIVATE EQUITY LLC us_ma 15	
2.	Edit cells Edit column Transpose Sort... View	pS DING ApS (35) dk 29	
3.	Trans	JNT L.P. (31) us_va 16	
4.	Reconcile	ainer Start reconciling... Facets Actions Copy reconciliation data... Use values as identifiers Add entity identifiers column Container Corp. (26) Container Corp. (26) CONTAINER CORP. (26)	us_co 15

A screen grab from OpenCorporates. Company names and numbers have been partially blurred.

Match each cell to its best candidate

- Create a new item for each cell
- Create one new item for similar cells
- Match all filtered cells to...
- Discard reconciliation judgments
- Clear reconciliation data

- Once you have manually selected all the matches or automatically selected the best matches, you'll see your matched list of companies:

10 rows

Show as: **rows** records Show: 5 10 25 50 rows

All	Name	Country code	Date active
1.	PRIVATE EQUITY LLC Choose new match	us_ma	15/05/2006
2.	HOLDING ApS Choose new match	dk	29/07/2020
3.	L.P. Choose new match	us_va	16/03/2015
4.	CONTAINER CORP. Choose new match	us_co	15/05/2008
5.	Holdings B.V. Choose new match	nl	15/05/2019
6.	LIMITED Choose new match	nz	14/01/2007
7.	GmbH Choose new match	de	16/05/2011
8.	LIMITED Choose new match	gb	16/05/2011
9.	INC. Choose new match	us_de	16/05/2007
10.	LIMITED Choose new match	us_tx	14/06/2020

A screen grab from OpenCorporates. Company names and numbers have been partially blurred.

Improve your Reconciliation: Add the Jurisdiction as a Service

If the companies in your list are all in one jurisdiction, you can constrain the reconciliation to only look for companies in that jurisdiction in the OpenCorporates database.

- To do this, add the country code to the URL you entered in the 'Add Standard Service' box:

Add standard service

Enter the service's URL:

A screen grab from OpenCorporates.

- Use the ISO3166 alpha codes. These are used throughout the IT industry by computer systems and software to ease the identification of country names. They

are also what OpenCorporates uses — you'll see them in the URLs on the website.

- Most jurisdictions are at the country level.

■

- The US, Canada and France however have jurisdictions at state/region level.
For those, put the country code_region/state code at the end of the url, e.g.
https://opencorporates.com/reconcile/us_mt for Montana.
- If you want to reconcile against the whole of the US, Canada or France, just put the country code at the end of the URL, e.g.
<https://opencorporates.com/reconcile/us>
- You can find all the jurisdiction codes by doing a search on the OpenCorporates website to see the code in the search result URL.

- Click 'Add Service'.
- You'll see that the jurisdiction-constrained reconciliation has been added on the left-hand side of the 'Reconciliation' box:

Reconcile column "Name"

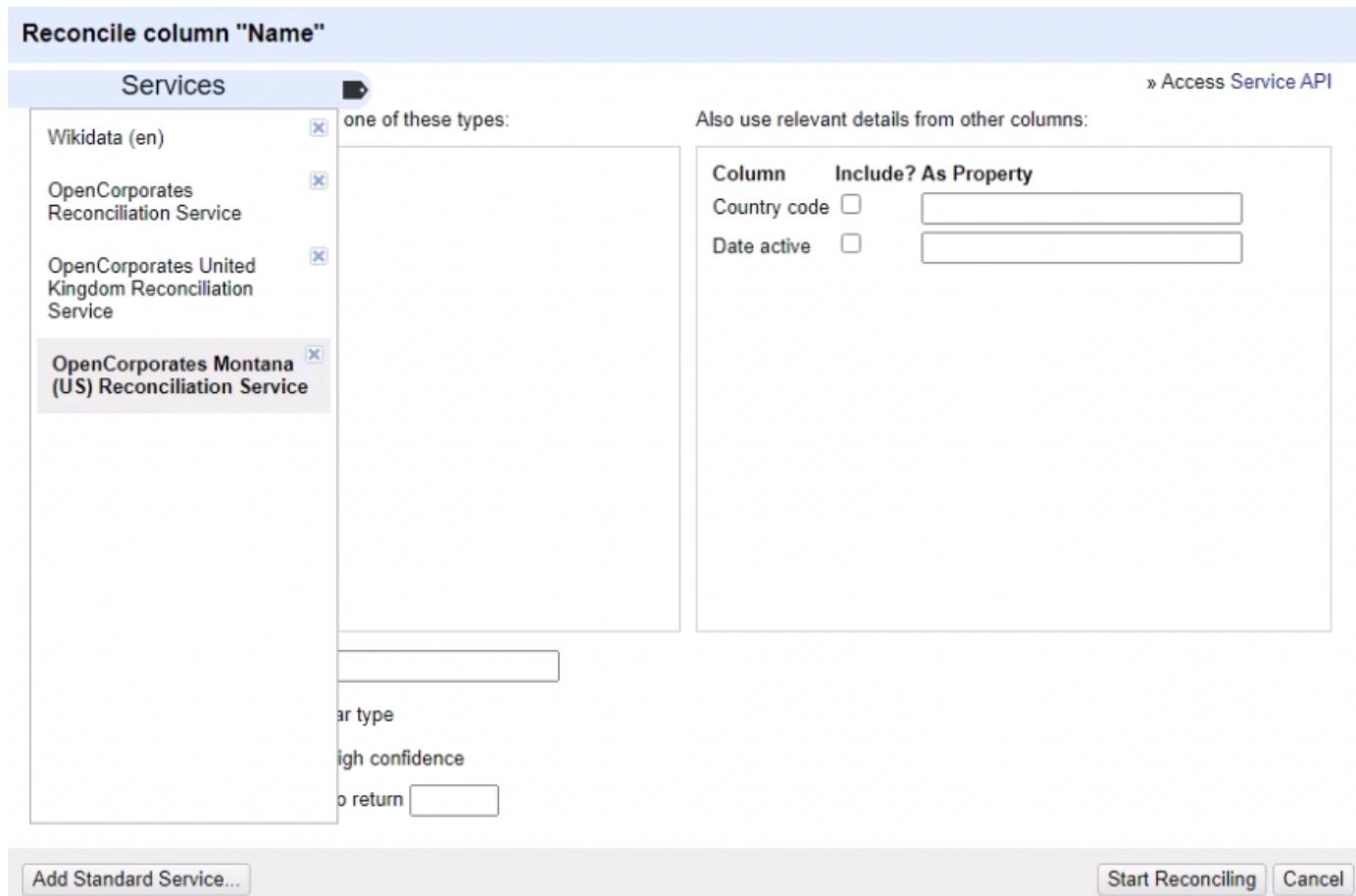
Services  [» Access Service API](#)

Column	Include? As Property
Country code	<input type="checkbox"/>
Date active	<input type="checkbox"/>

one of these types:
Wikidata (en) 
OpenCorporates Reconciliation Service 
OpenCorporates United Kingdom Reconciliation Service 
OpenCorporates Montana (US) Reconciliation Service 

Also use relevant details from other columns:

Add Standard Service... Start Reconciling Cancel



A screen grab from OpenCorporates.

In this image, two jurisdiction-constrained reconciliations have been added – one for the UK and one for the US state of Montana.

- In order to reconcile using one of the services you have added, you need to select it (so that has a pale grey background, like the Montana service in the image above), then click ‘Start Reconciling’

Improve your Reconciliation: Add Multiple Parameters

Adding multiple parameters to your reconciliation will improve the matching results. The reconciliation service can handle these two: country (jurisdiction) and date. You have to have those two columns in your CSV list.

- The date parameter is when a company was active or in existence. It helps the reconciliation to prioritise companies that were active on a specified date.
- To use the date parameter, you need to transform the date column in your uploaded CSV so that it is in a format that is recognised by the OpenCorporates API. To do this:
 - Start by clearing the reconciliation data (you need to do this as you are going to be doing a new reconciliation after you have done the transformation): Click the arrow next to ‘Name’ (or whatever the column header was in your data), then go to ‘Reconcile’, then ‘Actions’, then select ‘Clear reconciliation data’:

The screenshot shows the OpenRefine interface with a reconciliation task named "reconciliation_demo_NICAR csv". On the left, there are facets for "Name: judgment" and "Name: best candidate's score". The main area displays 10 rows of company data with columns: All, Name, Country code, and Date active. A context menu is open over the "Name" column, specifically over the first row. The menu path is "Name" > "Reconcile" > "Actions". The "Actions" submenu is expanded, showing options: "Start reconciling...", "Facets", "Actions" (which is selected and highlighted in blue), "Copy reconciliation data...", "Use values as identifiers", and "Add entity identifiers column". At the bottom of the submenu, the option "Clear reconciliation data" is also visible. The data table shows 10 rows of company information, with the first row being the focus of the reconciliation process.

A screen grab from OpenCorporates. Company names and numbers have been partially blurred.

To transform your date column, click on the arrow next to ‘Date Active’ (or whatever the column name was in your data), then ‘Edit cells’, then ‘Common transforms’, then select ‘To date’:

The screenshot shows a table with 10 rows of company data. The columns are labeled 'All', 'Name', 'Country code', and 'Date active'. The 'Date active' column contains dates like 'T4/06/2020'. A context menu is open over the last row of the 'Date active' column. The 'Edit cells' option under 'Common transforms' is selected, and the 'To date' option is highlighted in the dropdown menu.

All	Name	Country code	Date active
1.	Private Equity LLC	us_ma	
2.	HOLDING ApS	dk	
3.	L.P.	us_va	
4.	Container	us_co	
5.	Holdings	nl	
6.	Limited	nz	
7.	GmbH	de	
8.	Limited	gb	
9.	, Inc.	us_de	
10.	Enterprises	us_tx	T4/06/2020

A screen grab from OpenCorporates. Company names and numbers have been partially blurred.

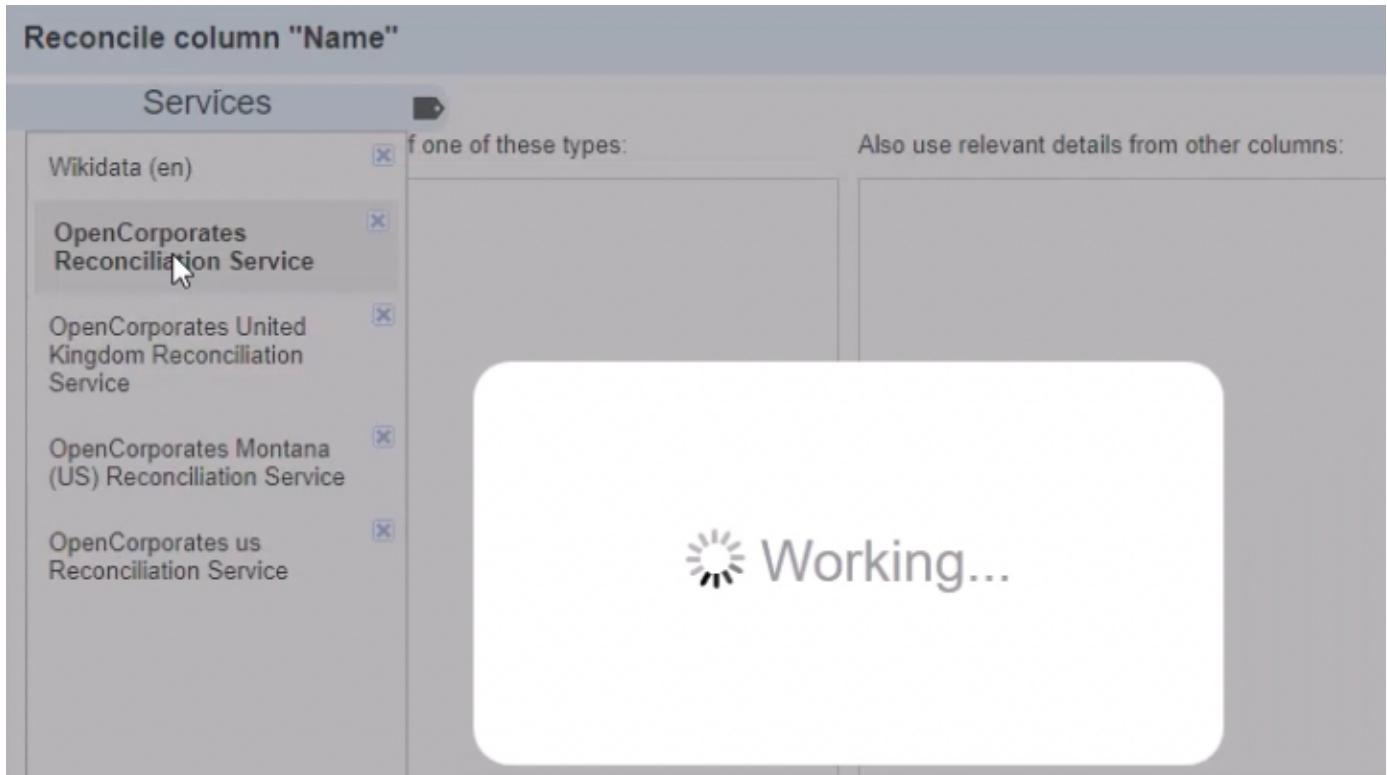
Your uploaded CSV will now look this this (don't worry about the 00:00 times):

The screenshot shows the same table after transformation. The 'Date active' column now displays dates in ISO 8601 format, such as '2006-05-15T00:00:00Z'. The other columns ('All', 'Name', 'Country code') remain the same.

All	Name	Country code	Date active
1.	Private Equity LLC	us_ma	2006-05-15T00:00:00Z
2.	HOLDING ApS	dk	2020-07-29T00:00:00Z
3.	L.P.	us_va	2015-03-16T00:00:00Z
4.	Container	us_co	2008-05-15T00:00:00Z
5.	Holdings	nl	2019-05-15T00:00:00Z
6.	Limited	nz	2007-01-14T00:00:00Z
7.	GmbH	de	2011-05-16T00:00:00Z
8.	Limited	gb	2011-05-16T00:00:00Z
9.	Inc.	us_de	2007-05-16T00:00:00Z
10.	Enterprises	us_tx	2020-06-14T00:00:00Z

A screen grab from OpenCorporates. Company names and numbers have been partially blurred.

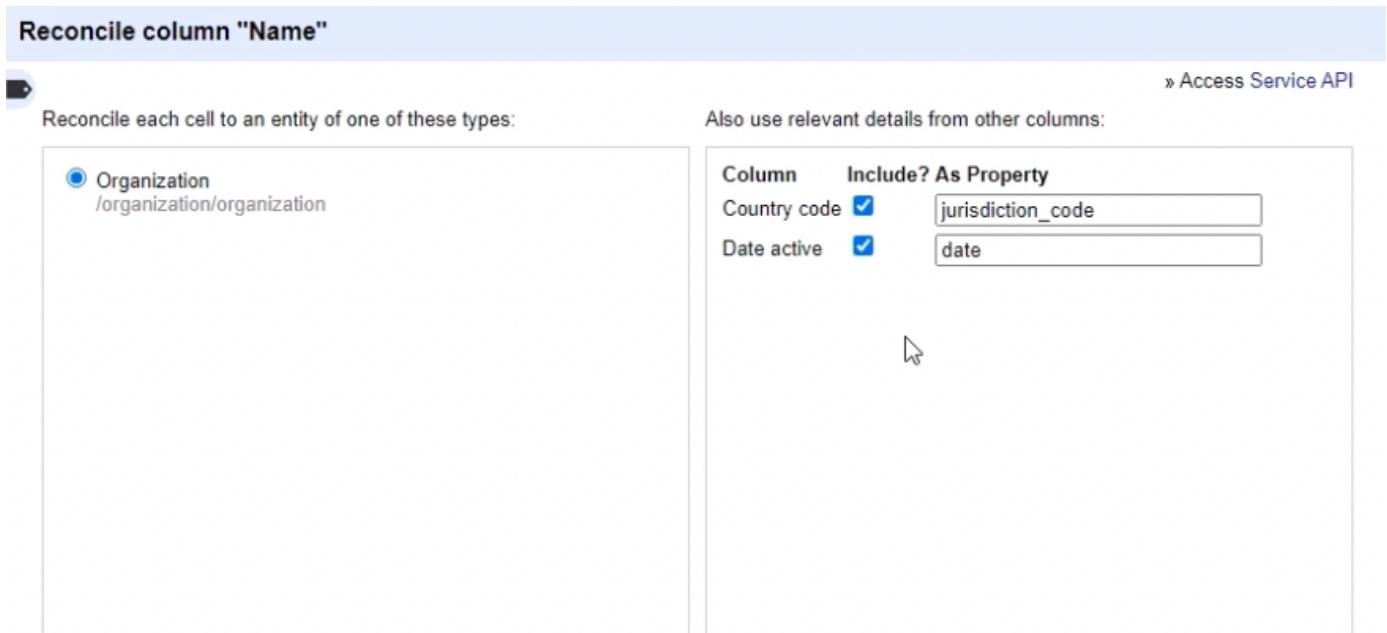
- Now you can select multiple parameters for reconciliation. First, click on the service that you want to use to reconcile:



A screen grab from OpenCorporates.

(In the case above, a non-jurisdiction specific service was selected.)

- Once the service has been selected, click the parameters you want on the right-hand side of the 'Reconcile' box:



A screen grab from OpenCorporates.

- Fill in the required parameter names in the 'As Property' text boxes (you have to use 'jurisdiction_code' and 'date' as these are the parameter names that the

OpenCorporates API recognises).

- If you use the 'date' parameter, a company will score higher if it was active at the date in your CSV, and will score a lot lower if it was not in existence or was inactive at that date.
 - This is particularly useful for jurisdictions that allow only one company with a particular name to be active at a point in time.
- Click 'Start Reconciling'
- If you have judgement and best candidate score boxes from previous reconciliations on the left-hand side, get rid of the old ones by clicking the 'x' at the top left of the boxes (the old ones are at the top), e.g.:

Facet / Filter Undo / Redo 3 / 3

Refresh Reset all Remove all

Name: judgment change
2 choices Sort by: name count
(blank) 1 none 10

Name: best candidate's score change reset
-20 — 100
7 0 4 0
✓ Numeric □ Non-numeric ✓ Blank □ Error

Name: judgment change
2 choices Sort by: name count
(blank) 1 none 10

Name: best candidate's score change reset
-20 — 100
7 0 4 0
✓ Numeric □ Non-numeric ✓ Blank □ Error

11 rows
Show as: rows

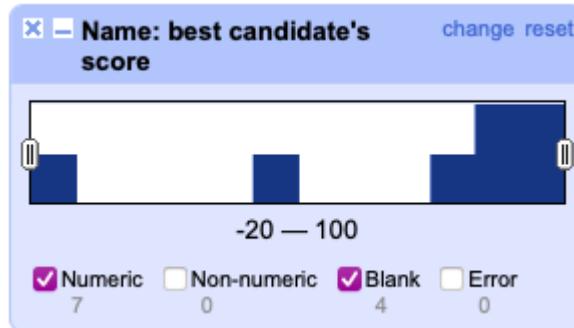
All N

1.	BOS	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
2.	H. SA	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
3.	GRE	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
4.	BALL	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
5.	Mara	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
6.	ACCI	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
7.	Maid	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
8.	REN	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
9.	MST	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
10.	HER	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>

A screen grab from OpenCorporates.

(Note: You need to do this to get the full results of the new reconciliation.)

- To filter by score, move the slider in this box:



A screen grab from OpenCorporates.

- Read more about the judgement and best candidate scores in the [OpenRefine documentation](#).
- Now that you have added the jurisdiction and date parameters to your reconciliation, you should have a lot fewer results to assess for correct matches.
- Before you can proceed, you need to manually accept matches, or use the 'Match each cell to its best candidate' function.

Pull data from OpenCorporates into OpenRefine

Now that you have your CSV list of companies matched to a company in the OpenCorporates database, you can pull in more data to give you a richer data set for your investigation. To do this:

- Start by pulling in the OpenCorporates URL for each company. To do this, click on the arrow next to 'Name' (or whatever you called that column), then 'Edit column', then select 'Add column based on this column':

The figure shows a context menu for the 'Name' column in OpenRefine. The menu items are: All, Facet, Text filter, Edit cells, Edit column (highlighted), Transpose, Sort..., View, Reconcile, OpenCorporates, Split into several columns..., Join columns..., Add column based on this column..., Add column by fetching URLs... (highlighted with a cursor), and Add columns from reconciled values... .

A screen grab from OpenCorporates.

In the 'Add column based on column Name' box that pops up, put this into the Expression text box: 'http://opencorporates.com'+cell.recon.match.id, and give your new column a name:

Add column based on column Name

New column name

On error set to blank store error copy value from original column

Expression

Language

'http://opencorporates.com'+cell.recon.match.id

No syntax errors

[Preview](#)

[History](#)

[Starred](#)

[Help](#)

row	value	URL
1.	Private Equity LLC	'http://opencorporates.com'+ce ...
2.	ApS	'http://opencorporates.com'+ce ...
3.	LLP	null
4.	Container	'http://opencorporates.com'+ce ...
5.	Holdings	'http://opencorporates.com'+ce ...
6.	Limited	null

row	value	URL
1.	Private Equity LLC	'http://opencorporates.com'+ce ...
2.	ApS	'http://opencorporates.com'+ce ...
3.	LLP	null
4.	Container	'http://opencorporates.com'+ce ...
5.	Holdings	'http://opencorporates.com'+ce ...
6.	Limited	null

A screen grab from OpenCorporates. Company names and numbers have been partially blurred.

Click 'OK'.

You should now see these columns:

10 rows

Show as: [rows](#) [records](#) Show: [5](#) [10](#) [25](#) [50](#) [100](#) [500](#) [1000](#) rows

All	Name	OpenCorporates	Country code	Date active
1.	PRIVATE EQUITY LLC Choose new match	'http://opencorporates.com/companies/us_ma/f...'	us_ma	2006-05-15T00:00:00Z
2.	ApS Choose new match	'http://opencorporates.com/companies/dk/f...'	dk	2020-07-29T00:00:00Z
3.	LLP <input checked="" type="checkbox"/> Create new item Search for match		is_va	2015-03-16T00:00:00Z
4.	CONTAINER CORP. Choose new match	'http://opencorporates.com/companies/us_co/f...'	us_co	2008-05-15T00:00:00Z
5.	Holdings B.V. Choose new match	'http://opencorporates.com/companies/nl/f...'	nl	2019-05-15T00:00:00Z
6.	Limited <input checked="" type="checkbox"/> Create new item		mz	2007-01-14T00:00:00Z

A screen grab from OpenCorporates. Company names and numbers have been partially blurred.

The OpenCorporates URLs are also a unique ID for each company, although they have been obscured in the image above.

- Next you need to pull all the data for each company through the API, and then parse it into separate columns.
- To pull all the data through the API, click on the arrow next to the 'Name', then 'Edit column', then select 'Add column by fetching URLs':

The screenshot shows a data grid with 10 rows. The 'Name' column is currently selected, indicated by a blue border around its header. A context menu is open over this column, specifically at the first row. The menu items are:

- Facet
- Text filter
- Edit cells
- Edit column** (this item is highlighted)
- Transpose
- Sort...
- View
- Reconcile

Under the 'Edit column' option, there is a secondary submenu with the following items:

- Split into several columns...
- Join columns...
- Add column based on this column...** (this item is highlighted)
- Add column by fetching URLs...
- Add columns from reconciled values...

A screen grab from OpenCorporates.

In the 'Add column by fetching URLs based on column Name' box that pops up, put this into the Expression text box: 'http://api.opencorporates.com' + cell.recon.match.id + '?api_token=XXXX', give your new column a name and enter 200 into the 'Throttle delay' box so that you don't overload the server:

Add column by fetching URLs based on column Name

New column name Throttle delay milliseconds

On error set to blank store error Cache responses

HTTP headers to be used when fetching URLs: [Show](#)

Formulate the URLs to fetch:

Expression Language [General Refine Expression Language \(GREL\)](#)

```
'https://api.opencorporates.com' + cell.recon.match.id + '?api_token=kxdLxCbK'
```

No syntax error.

Preview History Starred Help

row	value	URL
1.	Private Equity LLC	https://api.opencorporates.com/companies/us_ma/651 api_token=kxdLxCbK
2.	ApS	https://api.opencorporates.com/companies/dk/308208 api_token=kxdLxCbK
3.	, L.P.	https://api.opencorporates.com/companies/us_va/L009 api_token=kxdLxCbK
4.	Container	https://api.opencorporates.com/companies/us_co/1995 api_token=kxdLxCbK

OK Cancel

A screen grab from OpenCorporates. Company names and numbers have been partially blurred.

(The new column has been named 'OpenCorporates_JSON' as JSON is used to pull the data out, but you can call it what you want.)

Click 'OK'. It may take a little time for all the data to come down, especially if you have a long list.

You should end up with something that looks like this:

A screen grab from OpenCorporates. Company names and numbers have been partially blurred.

Parse the Fields that you Want

The data above is not usable and needs to be parsed into columns.

- To do that, click on the arrow next to ‘Name’, then ‘Edit column’, then select ‘Add column based on this column’:

All		Name	OpenCorporates_JSON
		1. PRIVATE EQUITY LLC Choose new match	Facet ▶ Text filter Edit cells ▶ Edit column ▶ Transpose ▶ Sort... View ▶ Add column based on this column... Add column by fetching URLs... Add columns from recognised values
		2. HOLDING ApS Choose new match	Text filter Edit cells ▶ Edit column ▶ Transpose ▶ Join columns... Add column based on this column... Add column by fetching URLs... Add columns from recognised values

A screen grab from OpenCorporates. Company names and numbers have been partially blurred.

- To parse the incorporation date, click on the arrow next to your new column (in this picture 'OpenCorporates JSON').

- In the 'Add column based on column OpenCorporates_JSON' box that pops up, put this into the Expression text box: value.parseJson()["results"]["company"]["incorporation_date"]
- Give your new column a name:

Add column based on column OpenCorporates_JSON

New column name	<input type="text" value="Incorporation_Date"/>						
On error	<input checked="" type="radio"/> set to blank <input type="radio"/> store error <input type="radio"/> copy value from original column						
Expression	Language <input type="button" value="General Refine Expression Language (GREL)"/> <code>value.parseJson()["results"]["company"]["incorporation_date"]</code> No syntax error.						
<input type="button" value="Preview"/> <input type="button" value="History"/> <input type="button" value="Starred"/> <input type="button" value="Help"/>							
<table border="1"> <thead> <tr> <th>row</th> <th>value</th> <th>value.parseJson()["results"]["..."]</th> </tr> </thead> <tbody> <tr> <td>1.</td> <td> <pre>{"api_version": "0.4.8", "results": {"company": {"name": "████████ PRIVATE EQUITY LLC", "company_number": "████████", "jurisdiction_code": "10-24", "dissolution_date": "2011-12-27", "company_type": "Foreign Limited Liability Company (LLC)", "registry_url": "http://corp.sec.state.ma.us/CorpWeb/FEIN=████████", "branch": "F", "branch_status": "branch of an out-of-jurisdiction company", "inactive": true, "current_status": null, "created_at": "2013-01-13T19:13:42+00:00", "updated_at": "2013-01-13T19:13:42+00:00"}}}</pre> </td> <td>2003-10-24</td> </tr> </tbody> </table>		row	value	value.parseJson()["results"]["..."]	1.	<pre>{"api_version": "0.4.8", "results": {"company": {"name": "████████ PRIVATE EQUITY LLC", "company_number": "████████", "jurisdiction_code": "10-24", "dissolution_date": "2011-12-27", "company_type": "Foreign Limited Liability Company (LLC)", "registry_url": "http://corp.sec.state.ma.us/CorpWeb/FEIN=████████", "branch": "F", "branch_status": "branch of an out-of-jurisdiction company", "inactive": true, "current_status": null, "created_at": "2013-01-13T19:13:42+00:00", "updated_at": "2013-01-13T19:13:42+00:00"}}}</pre>	2003-10-24
row	value	value.parseJson()["results"]["..."]					
1.	<pre>{"api_version": "0.4.8", "results": {"company": {"name": "████████ PRIVATE EQUITY LLC", "company_number": "████████", "jurisdiction_code": "10-24", "dissolution_date": "2011-12-27", "company_type": "Foreign Limited Liability Company (LLC)", "registry_url": "http://corp.sec.state.ma.us/CorpWeb/FEIN=████████", "branch": "F", "branch_status": "branch of an out-of-jurisdiction company", "inactive": true, "current_status": null, "created_at": "2013-01-13T19:13:42+00:00", "updated_at": "2013-01-13T19:13:42+00:00"}}}</pre>	2003-10-24					
<input type="button" value="OK"/> <input type="button" value="Cancel"/>							

A screen grab from OpenCorporates. Company names and numbers have been partially blurred.

- Click OK.
- Scroll to the right and you'll see the new incorporation date column:

	Incorporation_D
rpSummary.aspx?... lisher": "Massachusetts Secretary of the Commonwealth, A"], "alternative_names": [], "previous_names": ddress": "1... 0"], "ultimate_beneficial_owners": [], "filings": [], "officers": h": null}], {"officer": {"id": "...", "name": "... ", "position": "real ... er": "Det centrale HOLDING ApS", "start_date": "... ess": "Ny Kastelvej 2, ... 18T17:03:41+00:00", "updated_at": "2021-02- agent_address": {"street_address": "... Pennsylvania, "...", "industry_codes": owners": [], "filings": [], "officers": [{"officer": ": null}]}]} created_at": "2015-09-12T00:41:33+00:00", "updated_at": "2021- lawP111" = "mark_imanu_sulf null "register_name": "World	2003-10-24 2007-08-27 1993-03-22 1995-12-12

A screen grab from OpenCorporates. Company names and numbers have been partially blurred.

- To parse the current status, use this expression: value.parseJson()["results"]["company"]["current_status"]
- To parse other fields, look for their names in the JSON so that you can use their proper names in the expression, e.g.:
 -

({"name": "BALL METAL BEVERAGE CONTAINER CORP.", "company_number": "19951151557", "jurisdiction_code": "...

A screen grab from OpenCorporates.

In the above image you can see: name, company_number, jurisdiction_code.

Conclusion

It takes work to learn to use the OpenCorporates API, but it's worth it – you can save yourself days of manual labour. If you're going to be researching companies at scale, learn to use OpenRefine. It may be useful for other investigations — see the [list of](#)

reconciliation services that allow for data to be brought into OpenRefine (e.g. OpenSanctions).

This guide is built on the video 'CIJ Summer Workshops: OpenCorporates API for Beginners' on [this page](#).

The author would like to confirm that they have no financial or legal interest in OpenCorporates, nor prior personal relationship with its founders and employees. This article has been undertaken strictly from the perspective of an open source researcher.

Bellingcat is a non-profit and the ability to carry out our work is dependent on the kind support of individual donors. If you would like to support our work, you can do so [here](#). You can also subscribe to our Patreon channel [here](#). Subscribe to our [Newsletter](#) and follow us on Instagram [here](#), Twitter [here](#) and Mastodon [here](#).

Share this article



Katherine de Tolly

Katherine de Tolly is a Bellingcat Tech Fellow and is passionate about using bulk data processes to make companies and finance investigations easier.

Related articles



Corporate Finance

New Tools Dig Deeper Into Hard-to-Aggregate US Corporate Data



Finance Russia

Frontline Facilitators: How Secretive UK Partnerships Supply Wartime Russia