

Following the money using open-source information

Dataharvest

Le Monde

Maxime Ferrer

Data miner

@fneffer

ferrer@lemonde.fr

Jérémie Baruch

Data-driven reporter

@jeremie_baruch

baruch@lemonde.fr

Maxime Vaudano

Data-driven reporter

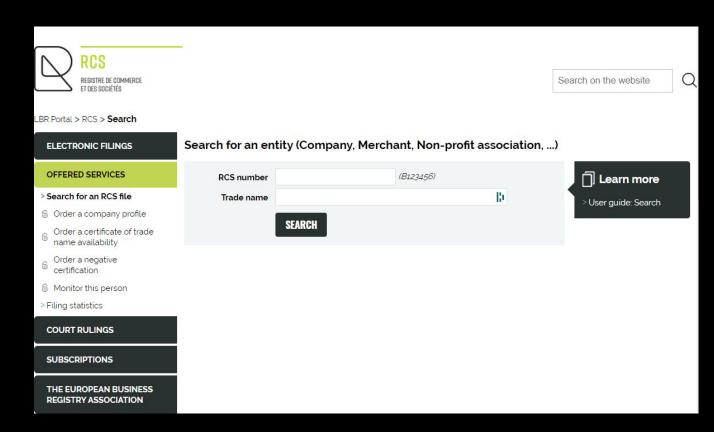
@mvaudano

vaudano@lemonde.fr

An old idea

2016

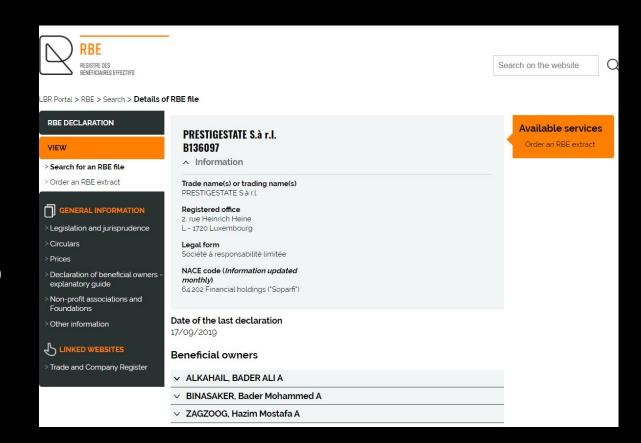
Luxembourg launches a new company register



The starting point

2019

Luxembourg is one of the first EU country to launch a beneficial ownership register



The challenges of scraping

```
try:
    html doc = RBE search(session, RCS)
    only this = SoupStrainer('div', id='content')
    soup = BeautifulSoup(html doc, 'lxml', parse only=only this)
    # 1: PRESENCE D'ERREURS
    if soup.find('div', class = 'msgErrorBox'):
        register_message = soup.find('div', class_='msgErrorBox').li.get_text()
        if ('any results' in register message): # si pas de résultat
            row['error_message'] = f'Company absent from the register : {register message} (998)'
            data.append(row)
        else:
            data.append(("error_message":f"Other error message : {register_message} (999)", "RCS_number":RCS, "fatal":True, "soup":soup.prettify()})
        return data
    # 2: INFORMATIONS GENERALES
        company details = soup.find("ul", class ="companyDetails")
        if company_details is None: # si pas de détails
                #row['trade_names_or_trading_names'] = soup.find('table', class_='commonTable').tbody.td.get_text(strip=True)
                list matches = soup.find("table", class ="commonTable")
                if list_matches is None: # Aucune info : probablement une erreur
                    data.append({"error message": "No info on the page (1000)", "RCS number": RCS, "fatal": True, "soup": soup.prettify()})
                else: # On trouve une liste : certificate of non-registration of BO ou autre
                    row['trade_names_or_trading_names'] = list_matches.tbody.find_all("tr")[0].find_all("td")[0].get_text(strip=True)
                    register_message = list_matches.tbody.find_all("tr")[0].find_all("td")[2].get_text(strip=True)
```

- Overcome the technical hurdles
- Retro-engineer the platform to understand the data
- Keep track of the changes
- Export the data in a reporter-friendly format

Financial data extraction

D.	Cui	Current assets		1151	151	6.986.089,00	152	69.119.793,00
	I.	Stocks		1153	153		154	
		1.	Raw materials and consumables	1155	155		156	
		2.	Work in progress	1157	157		158	
		3.	Finished goods and goods for resale	1159	159		160	
		4.	Payments on account	1161	161		162	
	II.	Debtors		1163		6.208.298,00	164	68.639.146,00
		1.	Trade debtors	1165	165		166	
			becoming due and payable within one year	1167	167		168	
			b) becoming due and payable after more than one year	1169	169			
		2.	Amounts owed by affiliated undertakings	1171	171	6.195.988,00	172	68.638.070,00
			becoming due and payable within one year	1173	173	6.195.988,00	174	68.638.070,00
			 b) becoming due and payable after more than one year 	1175	175		176	
		3.	Amounts owed by undertakings with which the undertaking is linked by virtue of participating interests	1177	177		178	
			becoming due and payable within one year	III I	179			
			b) becoming due and payable after more than one year	1181	181		182	
		4.	Other debtors	1183			184	1.076,00

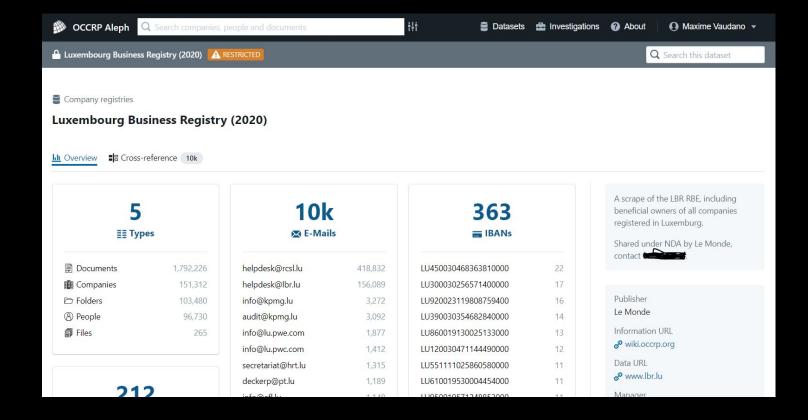
 Regex (regular expressions) to extract formatted accounting sheets

Current assets .*\s+151\s\s+(-?[\d+\.?]+[,\d+]+)\s+152\s+(-?[\d+\.?]+[,\d+]+)?

How to organise and search the data



(OCCRP)



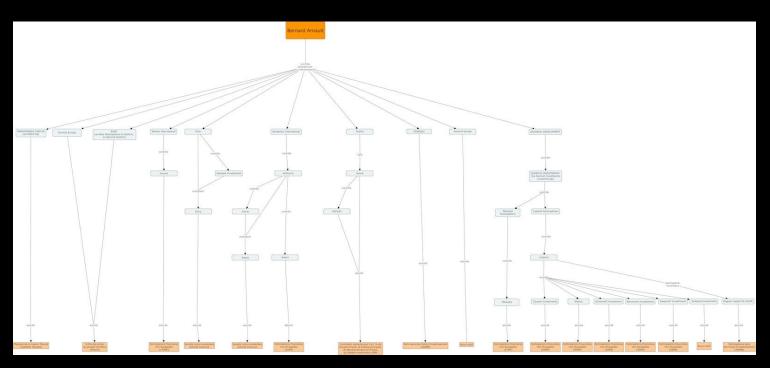
How to organise and search the data

Spreadsheets

▼ company_rcs	▼ company_name	company_address	company_creati	company_deleti	company_nace_code	company_legal	company_curre	person_name	person_birthda	person_birthda	t person_birth_c
B306	SOCIETE MARIA RHEINSHEIM	5, Av. Marie Thérèse L - 2132 Luxembourg	1937-02-05		68.200 Renting and operating of own or leased real estate	Société anonyme	Administrators listed by default	WAGENER, Marc Joseph Emile	1965-08-26	26/08/1965	Luxembourg
B306	SOCIETE MARIA RHEINSHEIM	5, Av. Marie Thérèse L - 2132 Luxembourg	1937-02-05		68.200 Renting and operating of own or leased real estate	Société anonyme	Administrators listed by default	HENTGEN, Pierre Aloyse	1961-11-08	08/11/1961	Luxembourg
B306	SOCIETE MARIA RHEINSHEIM	5, Av. Marie Thérèse L - 2132 Luxembourg	1937-02-05		68.200 Renting and operating of own or leased real estate	Société anonyme	Administrators listed by default	VANOLST, Jean Joseph Hortense	1944-02-02	02/02/1944	Luxembourg
B306	SOCIETE MARIA RHEINSHEIM	5, Av. Marie Thérèse L - 2132 Luxembourg	1937-02-05		68.200 Renting and operating of own or leased real estate	Société anonyme	Administrators listed by default	SCHARES, Nicolas	1947-05-15	15/05/1947	Ettelbruck
B306	SOCIETE MARIA RHEINSHEIM	5, Av. Marie Thérèse L - 2132 Luxembourg	1937-02-05		68.200 Renting and operating of own or leased real estate	Société anonyme	Administrators listed by default	WAGENER, Léon Nicolas	1962-04-12	12/04/1962	Ettelbruck
	Prost Group S.A.	Zone Industrielle Rolach L - 5280 Sandweiler	1925-02-07		46.460 Wholesale of pharmaceutical goods	Société anonyme	Administrators listed by default		1960-10-22	22/10/1960	Luxembourg
	Société de Transport d'Energie Electrique du Grand-Duché de Luxembourg	4, Rue de Soleuvre L - 4321 Esch-sur-Alzette	1927-12-13		35.140 Trade of electricity	Société coopérative	Confidentiality clause	BASTIAN, Roland Alphonse Charles	1962-12-13	13/12/1962	Luxembourg
	Société de Transport d'Energie Electrique du Grand-Duché de Luxembourg	4, Rue de Soleuvre L - 4321 Esch-sur-Alzette	1927-12-13		35.140 Trade of electricity	Société coopérative	Confidentiality clause	KOEPP, Carlo Roger Lucien Guillaume Jacques	1956-08-16	16/08/1956	Luxembourg
	Société de Transport d'Energie Electrique du Grand-Duché de Luxembourg	4, Rue de Soleuvre L - 4321 Esch-sur-Alzette	1927-12-13		35.140 Trade of electricity	Société coopérative	Confidentiality clause	SLIEPEN, Christian	1978-06-05	05/06/1978	Luxembourg
	Société de Transport d'Energie Electrique du Grand-Duché de Luxembourg	4, Rue de Soleuvre L - 4321 Esch-sur-Alzette	1927-12-13		35.140 Trade of electricity	Société coopérative	Confidentiality clause	BAULESCH, Patrice Mathias Jean-Pierre	1964-01-27	27/01/1964	Esch-Sur-Alzette
	Société de Transport d'Energie Electrique du Grand-Duché de Luxembourg	4, Rue de Soleuvre L - 4321 Esch-sur-Alzette	1927-12-13		35.140 Trade of electricity	Société coopérative	Confidentiality clause	RODRIGUEZ, Emmanuel Alain Vincent	1972-12-28	28/12/1972	Chalons En Champagne
	Société de Transport d'Energie Electrique du Grand-Duché de Luxembourg	4, Rue de Soleuvre L - 4321 Esch-sur-Alzette	1927-12-13		35.140 Trade of electricity	Société coopérative	Confidentiality clause	DEBLAYE, Nicolas Hubert Olivier	1975-06-12	12/06/1975	Bourg La Reine
B1020	FERROKNEPPER BUDERUS S.A.	Zone Industrielle Esch-Schifflange L - 4149 Esch-sur-Alzette	1929-02-02		46.741 Wholesale of hardware	Société anonyme	Administrators listed by default	THIEL, STEFAN PETER	1972-04-26	26/04/1972	Tettnang
B1020	FERROKNEPPER BUDERUS S.A.	Zone Industrielle Esch-Schifflange L - 4149 Esch-sur-Alzette	1929-02-02		46.741 Wholesale of hardware	Société anonyme	Administrators listed by default	WEBER GEBOREN MÜLLER, STEFAN	1966-11-04	04/11/1966	Gottingen
B1020	FERROKNEPPER BUDERUS S.A.	Zone Industrielle Esch-Schifflange L - 4149 Esch-sur-Alzette	1929-02-02		46.741 Wholesale of hardware	Société anonyme	Administrators listed by default	HEINZ, Charles Joseph	1941-12-20	20/12/1941	Luxembourg
B1363	Sofina Capital S.A.	12, rue Léon Laval L - 3372 Leudelange	1931-05-12		64.202 Financial holdings ("Soparfi")	Société anonyme	Administrators listed by default	GILLET, François	1960-02-06	06/02/1960	Anvers
B1428	Société Luxembourgeoise ETAM	35, Avenue de la Gare L - 1611 Luxembourg	1931-09-21		47.710 Retail sale of clothing in specialised stores	Société anonyme	UBO listed	MILCHIOR, PIERRE	1933-11-26	26/11/1933	Neuilly-Sur-Seine
B1428	Société Luxembourgeoise ETAM	35, Avenue de la Gare L - 1611 Luxembourg	1931-09-21		47.710 Retail sale of clothing in specialised stores	Société anonyme	UBO listed	AMATO EPOUSE MILCHIOR, MICHELE RACHEL	1951-04-14	14/04/1951	Neuilly-Sur-Seine
B1428	Société Luxembourgeoise ETAM	35, Avenue de la Gare L - 1611 Luxembourg	1931-09-21		47.710 Retail sale of clothing in specialised stores	Société anonyme	UBO listed	MILCHIOR, LAURENT	1974-11-22	22/11/1974	Neuilly-Sur-Seine
B1428	Société Luxembourgeoise ETAM	35, Avenue de la Gare L - 1611 Luxembourg	1931-09-21		47.710 Retail sale of clothing in specialised stores	Société anonyme	UBO listed	MILCHIOR, VANESSA	1977-01-05	05/01/1977	Paris 16
B1860	ENGIE Invest International S.A.	65, Avenue de la Gare L - 1611 Luxembourg	1933-11-10		64.202 Financial holdings ("Soparfi")	Société anonyme	Administrators listed by default	VAN DEN HOVE, Thierry	1970-02-25	25/02/1970	Turnhout
B1860	ENGIE Invest International S.A.	65, Avenue de la Gare L - 1611 Luxembourg	1933-11-10		64.202 Financial holdings ("Soparfi")	Société anonyme	Administrators listed by default	SANTELLI, Shepherd Sansonise	1975-03-09	09/03/1975	Port-Au-Prince
B1860	ENGIE Invest International S.A.	65, Avenue de la Gare L - 1611 Luxembourg	1933-11-10		64.202 Financial holdings ("Soparfi")	Société anonyme	Administrators listed by default	GUIOLLOT, Pierre	1968-04-21	21/04/1968	Marseille
B2017	IMMOBILIERE François MULLER	8, rue d'Orange L - 2267 Luxembourg	1934-10-01		45.112 Retail trade of cars and light motor vehicles	Société à responsabilité limitée	UBO listed	KERSCHEN, André Justin Charles	1946-10-18	18/10/1946	Luxembourg
B2047	Société Immobilière du Golf	1, route de Trèves L - 2633 Senningerberg	1934-12-22		68.200 Renting and operating of own or leased real estate	Société anonyme	Administrators listed by default	CLASEN, Bernard Joseph Max	1947-04-19	19/04/1947	Luxembourg
B2047	Société Immobilière du Golf	1, route de Trèves L - 2633 Senningerberg	1934-12-22		68.200 Renting and operating of own or leased real estate	Société anonyme	Administrators listed by default	CLASEN, Antoine André	1982-12-16	16/12/1982	Luxembourg

How to organise, search and understand the data

Mindmapping tools



How to find stories

100 k 20 stories

How to find stories

Batch searches

with predefined lists

Cross references

with other datasets

Financial accounts analysis

Specific data extractions

Human verification

on names

How to find stories

Multiple criminals opened companies in Luxembourg without raising red flags

A major loophole for avoiding tax on French real estate was exploited for years

A large French luxury group paid offshore wages to its top executives

A famous chef shifted his royalties to Luxembourg to avoid taxes

Most large French
groups have
offshore
subsidiaries in
Luxembourg

Several Italians with ties to the 'Ndrangheta have opened weird Luxembourg companies

How to build meaningful figures

6.5 trillion €

The (minimal) amount of assets held by Luxembourg offshore companies

45%

The share of offshore structures in the Luxembourg companies

1 in 2

The share of Luxembourg companies failing to comply with the beneficial ownership mandatory declaration

266

The number of Forbes list billionaires appearing in the data

37/50

The number of French richest families appearing in the data

1804

The record number of companies registered at the same address (6 rue E. Rupert)

7 main partners



≈ 10 topic-by-topic partners







•••

Second round partners



Communication

Signal messenger

Shared notes

Atlassian Confluence (via OCCRP)

Data sharing

OCCRP Aleph

How to collaborate with NGOs and academics





Après l'enquête OpenLux, l'appel de quatorze économistes et personnalités à « créer d'urgence un cadastre financier européen »

Seule une petite partie des patrimoines financiers étant couverte par les registres existants, l'Union européenne doit s'attaquer à cette opacité, préconise le collectif dans un appel lancé à l'initiative de l'économiste Gabriel Zucman.

Publié le 09 mars 2021 à 17h20 - Mis à jour le 10 mars 2021 à 06h48 | 🐧 Lecture 3 min.



ÉVITEMENT FISCAL : LE RÔLE TOUJOURS PLUS IMPORTANT DU LUXEMBOURG

IIBLIÉ LE 14 EÉVRIER 2021

Un article d'Amélie Guillin, Vincent Bouvatier, Université Paris-Est Créteil (UPEC), Gunther Capelle-Blancard, Université Paris 1 Panthéon-Sorbonne, Anne-Laure Delatte, Université Paris Dauphine - PSL et Vincent Vicard (CBII)



Tutorial How to find information on a Luxembourg company

- 1 Identify the RCS number
- 2 Lookup for the beneficial owners in the RBE
- 3 Lookup for corporate documents (shareholders, administrators, addresses) in the RCS
- 4 Lookup for the assets (real estate, subsidiaries) in the annual accounts
- 5 Lookup for the financial flows

Case study: Gainsboro Developments (B198410)

Tutorial How to find information on a Luxembourg company

Other useful sources:

- opencorporates.com
- opengazettes.com
- register.openownership.org
- opencorporation.org
- lei.report
- reportfinder.eu
- List of EU business registers (bit.ly/EUBusinessRegisters)
- WIPO Global Brand Database (wipo.int/branddb/)