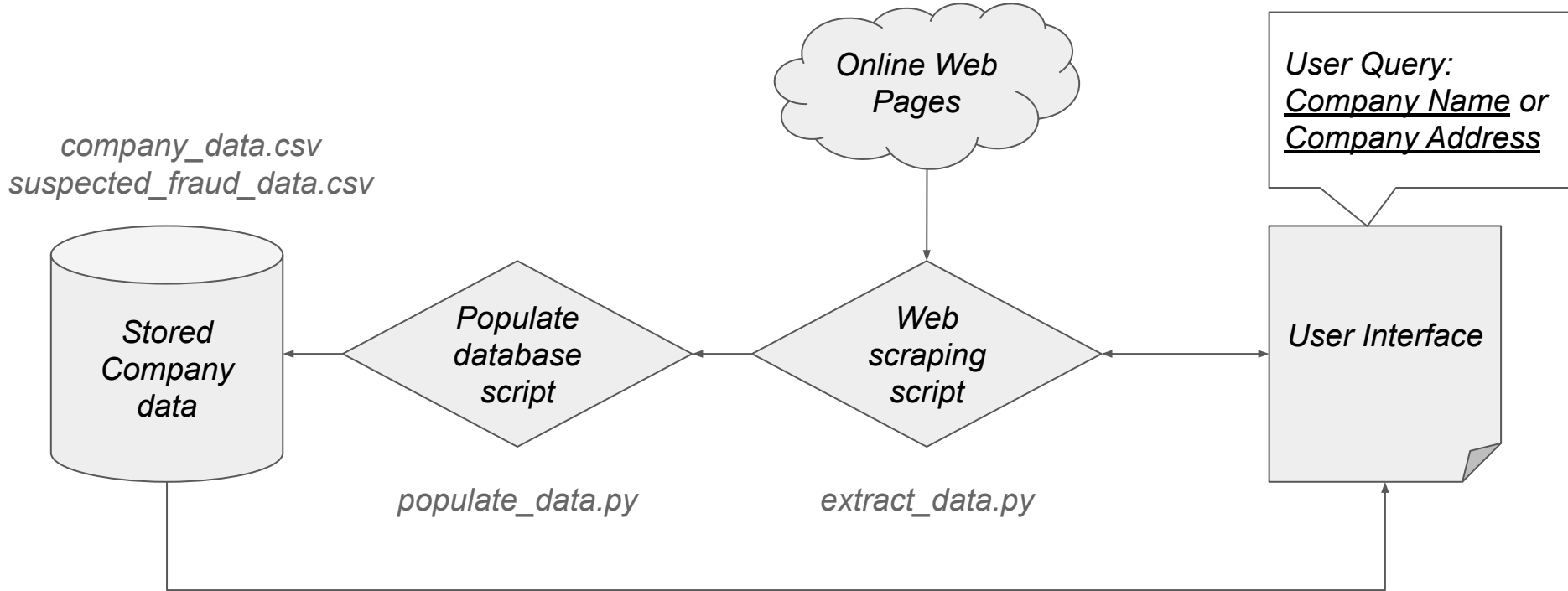




2021 ESDC HACKATHON TEAM ICE CREAM!!

So creamy

System Overview

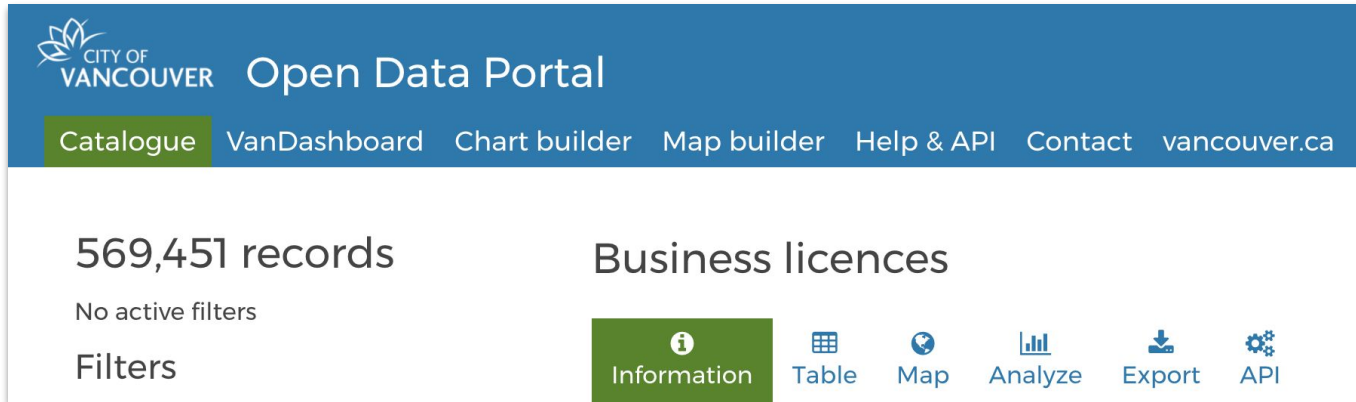


Data Scraping

- **Three main data sources** used
 - Two for company information
 - One for red flag detection
- All data sources extracted **follow** the websites' **terms and conditions**
 - Some sites that we would have liked to use otherwise specifically state in their terms of use that web scraping is not permitted, such as <https://beta.canadasbusinessregistries.ca>
- Variety of scraping methods used to extract data including **Selenium**, **BeautifulSoup**, **GET Requests**
- Data formats parsed included **JSON** and **XML**
- **Missing data** is displayed as “Unknown” in the application

Data Source 1: City of Vancouver Open Data Portal

- Used for populating **business data fields**
- The sourced data is updated regularly, previous folder years are expected to remain static
- Accessed via **public API**



The screenshot shows the City of Vancouver Open Data Portal. The header is blue with the City of Vancouver logo and the text 'Open Data Portal'. Below the header is a navigation bar with links: 'Catalogue' (highlighted in green), 'VanDashboard', 'Chart builder', 'Map builder', 'Help & API', 'Contact', and 'vancouver.ca'. The main content area is white and displays '569,451 records' and 'Business licences'. Below the record count is the text 'No active filters' and a 'Filters' link. A row of interactive buttons is shown: 'Information' (highlighted in green), 'Table', 'Map', 'Analyze', 'Export', and 'API'.

CITY OF VANCOUVER Open Data Portal

Catalogue VanDashboard Chart builder Map builder Help & API Contact vancouver.ca

569,451 records

No active filters

Filters

Information Table Map Analyze Export API

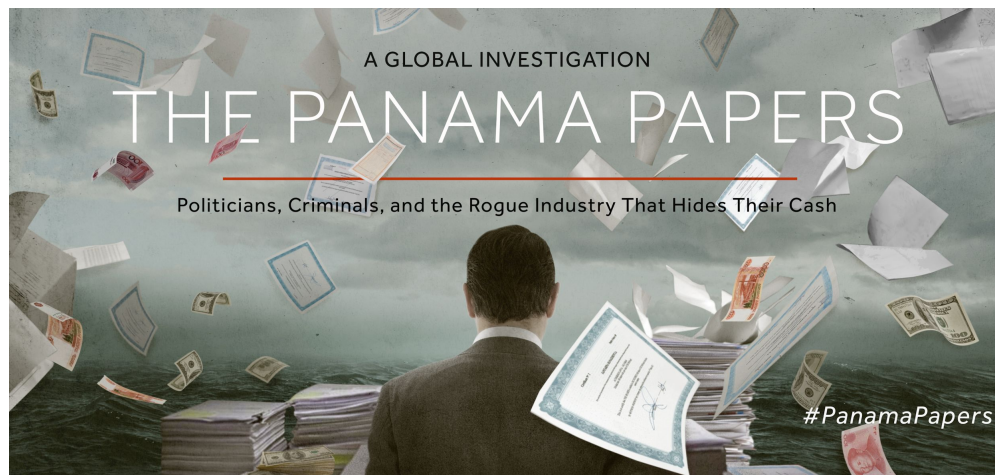
Data Source 2: Federal BN Registry

- Used for populating **business data fields**
- The data is updated weekly according to comments on the site
- One script **automatically downloads** the data in bulk as **XML** files
- Another script scraps the director names based on extracted company names and IDs using **Selenium**



Data Source 3: Offshore leaks database

- Used for **legal check** and **offshore accounts** for potential fraud detection
- Multiple sub-sources scraped, including:
 - Offshore leaks
 - Bahamas Leaks
 - Panama Papers
 - [CanLII Federal](#)
 - [CanLII Provincial \(BC\)](#)



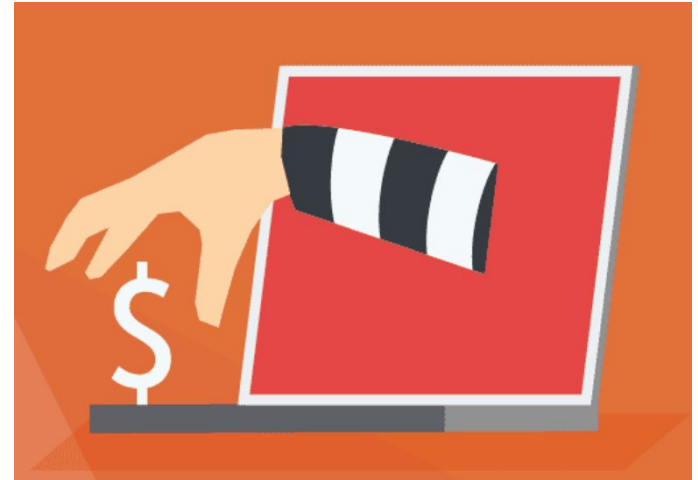
Extracted Data and Storage

- Data details
 - Over 1,000,000 rows of company details exported
 - ~790,000 unique corporation IDs
 - New fields include
 - Business numbers
 - Addresses for federally registered companies
 - Aliases
 - Directors
- Storage system
 - Currently using a series of csv files
 - Future plan is to migrate to database



Data analysis

- Fraud detection algorithms:
 - Word similarity comparisons (Levenshtein distance)
 - Business number validation
- Fraud detection reasoning:
 - Legal history with regards to bad faith
 - Multiple addresses for the same building
 - Lack of electronic paper trail
 - Mismatched addresses



User Interface

- UI is built with [Streamlit](#)
- Designed with the end user in mind
 - Able to search with either
 - Company address
 - Company name
 - Company name AND address
 - Easy to quickly read summary
 - Side-by-side maps for comparison
 - Displays key information
 - Includes helpful links for manual review
 - Helpful error messages
 - Fraud/Legal checks with suspicion score

Enter an address, business name, or both!

Enter an address

Enter a business name

meat and bread

Search

Results for Meat And Bread

Company Information

Name: **Meat And Bread**

Address: **Unit 796 7 W BROADWAY Vancouver BC CA V5Z 1H2**

Phone: **Unknown**

Number of employees: **0**

Estimated years active: **Unknown**

Director names: **Unknown**

Search Links

[Exact match for Meat And Bread](#)

[Partial match for Meat And Bread](#)

[Search for Meat And Bread](#)

[Search for Meat And Bread](#)

[Exact match for Unit 796 7 W BROADWAY Vancouver BC CA V5Z 1H2](#)

[Partial match for Unit 796 7 W BROADWAY Vancouver BC CA V5Z 1H2](#)

[Search for Unit 796 7 W BROADWAY Vancouver BC CA V5Z 1H2](#)

[Search for Unit 796 7 W BROADWAY Vancouver BC CA V5Z 1H2](#)

Other Matching Results

BusinessName	BusinessTradeName	Ad
		empty

Offshore Leaks Database Check

No red flag found for Unit 796 7 W BROADWAY Vancouver BC CA V5Z 1H2

	country_codes
8242 Elliott Street, V.	CAN
310-1755 WEST BROADWAY.	CAN
2360 Oliver Crescent, ..	CAN
1962 MILLS ROAD SIDNEY.	CAN
#200 - 1465 ELLIS STRE.	CAN

No red flag found for Meat And Bread

	country_codes
Sonoza Diamond Corpora..	CAN
Lenka Trust	CAN
Glencoe ESP (Canada) ..	CAN
Managed Care	CYM
Malta Convention Buses.	MLT
Meryland Limited	nan
East Wind Limited	CYM
BetaCat Fund Ltd.	BMU

CanLii Legal Check

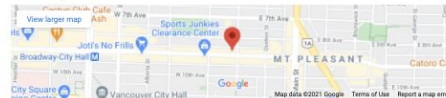
No red flag found for Meat And Bread

	court
Matute Andrade	fst https://www
Marchant Andrade	fst https://www
Bazge Andrea	fst https://www
Hexera Andrade	fst https://www
Matsqui Indian Band	fst https://www

Fraud Check

No fraud detected

Google maps search by address (Unit 796 7 W BROADWAY Vancouver BC CA V5Z 1H2)



Google maps search by business name (Meat And Bread)



View Source Data

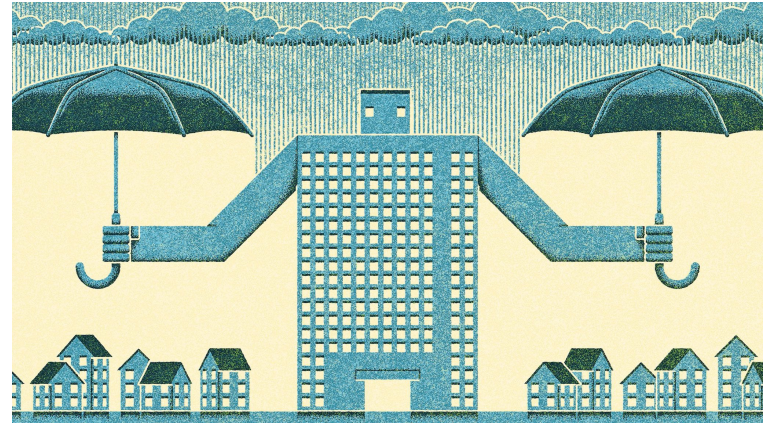


Usage Instructions

- Install by following instructions on <https://github.com/sbabicki/ice-cream>
- Enter in **business name**, **address**, or **both**
- If there are multiple matches for the same address these matches will also be shown along with key characteristics
- Matches for related companies and individuals (the latter yet to be implemented) that have a higher likelihood of bad faith actions

Ethical Implications

- Bias concerns, the following minorities may be targeted disproportionately:
 - **Immigrants** - may not be familiar with Canadian regulations
 - **Technologically illiterate** - lower online presence
 - **Lack of resources** - not enough resources to pay to advertise or set up online presence
 - **Type of business** - certain types businesses will have less of an online presence
 - **Legal records** - relying heavily on past behaviour to predict future behaviour
- Potential ethical concerns for our product:
 - Providing legal records
 - May not otherwise be checked
 - Maps displayed for all companies
 - May lead to subconscious biases



Future plans

- Additional Feature Aspirations
 - Move data sources to database
 - Additional data sources
 - Partial string matching for address and business name lookup
 - Fuzzy matching improvement for Offshore Leaks DB
 - Search by director name
 - Automated fraud detection score based on company characteristics
 - Link for viewing matching data result
 - API integration with app
- We are confident that with slightly more time we could implement many of these features!

Questions for us?

