



Connecting the Dots: From Gazette Automation to Probabilistic Entity Resolution

R Exchange 2026

Jane Li & Nilani Algiriyage
Data Science Team | Joint Targeting and Analytics
Immigration New Zealand

New Zealand Companies Office (NZCO)

- All companies operating in NZ must be **registered with NZCO**
- Incorporating on the **Companies Register** under the Companies Act 1993
- Registration provides a **NZBN** and enables legal operation and compliance
- Holds information on **registered directors** and shareholders



Immigration New Zealand (INZ)

- To **hire migrant** workers, employers must obtain **Employer Accreditation (EA)**
- Mandatory under the **Accredited Employer Work Visa (AEWV)** system
- Migrants' work visas are tied to the relevant EA



NZ Gazette

- **Official Government newspaper** and authoritative journal of constitutional record
- Various notices including those related to **liquidations**



Problem

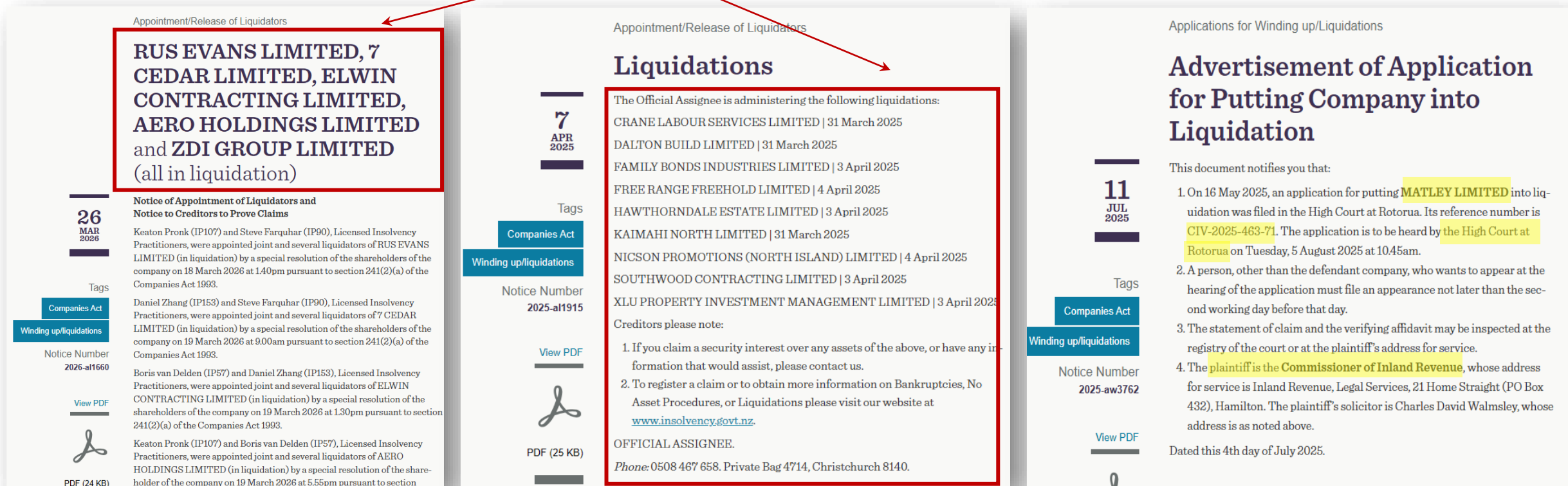
What happens when a company is announced for liquidation / receivership?

- Companies are only marked “Removed” in NZCO after official deregistration.
- INZ receives NZCO updates monthly, after removal has already occurred.
- There is a significant lag between liquidation decision and official removal.
- INZ requires this information as early as possible:
 - To revoke accreditation
 - Place necessary alert and warnings
 - Work with affected migrants
- Data doesn't align:

Gazette	<ul style="list-style-type: none">• No NZBN	}	<ul style="list-style-type: none">• <i>Mismatched IDs</i>• <i>Lack of ID</i>• <i>Siloed data</i>• <i>Different formats</i>
NZCO	<ul style="list-style-type: none">• NZBN• No director id		
INZ	<ul style="list-style-type: none">• NZBN• Application ID• Client ID		

Companies listed in the notice title vs body

Additional info we need for custom alert messages



The image displays three screenshots of Gazette notices. The first screenshot shows a notice titled "Appointment/Release of Liquidators" with a red box around the title listing "RUS EVANS LIMITED, 7 CEDAR LIMITED, ELWIN CONTRACTING LIMITED, AERO HOLDINGS LIMITED and ZDI GROUP LIMITED (all in liquidation)". The body text lists "RUS EVANS LIMITED" and "7 CEDAR LIMITED" but omits "ELWIN CONTRACTING LIMITED" and "ZDI GROUP LIMITED". The second screenshot shows a notice titled "Liquidations" with a red box around the title. The body text lists several companies including "CRANE LABOUR SERVICES LIMITED", "DALTON BUILD LIMITED", "FAMILY BONDS INDUSTRIES LIMITED", "FREE RANGE FREEHOLD LIMITED", "HAWTHORNDALE ESTATE LIMITED", "KAIMAHI NORTH LIMITED", "NICSON PROMOTIONS (NORTH ISLAND) LIMITED", "SOUTHWOOD CONTRACTING LIMITED", and "XLU PROPERTY INVESTMENT MANAGEMENT LIMITED". The third screenshot shows a notice titled "Applications for Winding up/Liquidations" with a red box around the title "Advertisement of Application for Putting Company into Liquidation". The body text mentions "MATLEY LIMITED" and "the High Court at Rotorua" but omits "the High Court at Rotorua" and "5 August 2025 at 10.45am".

➤ Gazette is a valuable signal, but it's completely *name-based* and *largely unstructured*.

Challenges with NZ Gazette

- Notices do not contain companies' **NZBN**
- Notices come in **various format**
- May include many companies per notice
- No fixed schedule for publications

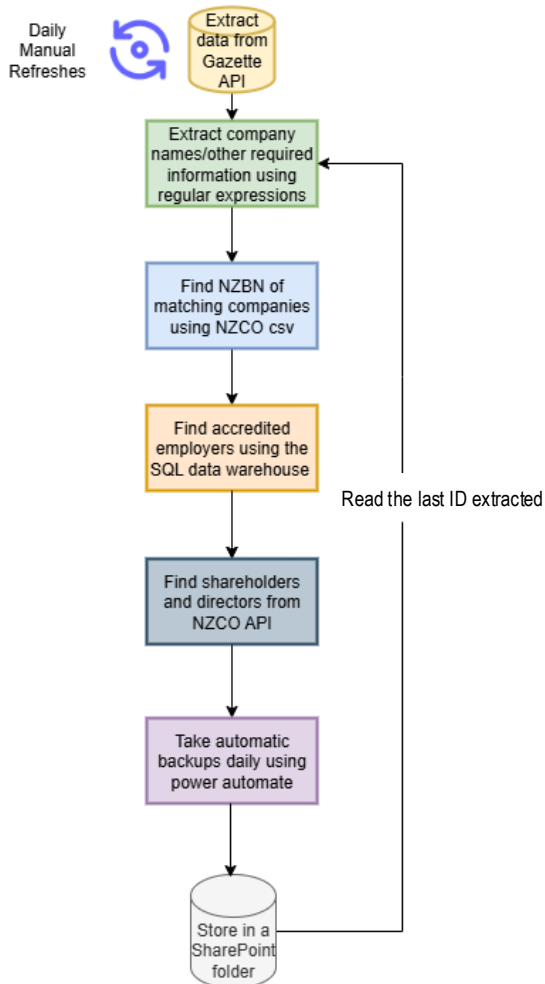
Solution to *automatically*

- ✓ Crawl notices via the Gazette API
- ✓ Extract company names
- ✓ Match names to NZBNs using NZCO data
- ✓ Cross-reference INZ applications
- ✓ Generate custom alert messages
- ✓ Maintain a database of daily extractions

- Automating Gazette monitoring **reduced manual effort** and **improved early risk detection**, and
- Exposed a **deeper problem: resolving entities** across systems **without unique identifiers** — solved through **probabilistic matching**.

Previous workflow

- Monitor Gazette for new notices
 - Manually extract names of companies
 - Search their NZBNs in the NZCO register
 - Cross-check NZBNs with INZ records
 - If found, create custom alert messages
- *Tedious, time-consuming, fully manual, 1FTE*



➤ Capturing Required Data using APIs

- Application Programming Interface (API)
- Gazette API: <https://digitalnz.org/developers/>
- NZCO API: <https://portal.api.business.govt.nz/>

Example:

- We send a request to an API (e.g., "Give me notices with keywords "liquidation" from today").
- Gazette API fetches the data from the source and returns it to our dashboard.

➤ Automatic text extraction

- Regular Expression (RegEx)
- A sequence of characters that specifies a match pattern in text.

Example:

Want to match

Any phone number in this format:
123-456-7890

RegEx expression

`\d{3}-\d{3}-\d{4}`

Explanation

- \d means "a digit" (0–9)
- {3} means "exactly 3 digits"
- - a dash

➤ Technologies used



Gazette Extract

Matched EAs

Last Extracted Date: 2/03/2026

Notice Number

All

Notice Display Date

All

NZBN

All

Entity Name

All

Clear all filters

Relative Extracted Date

Last 1 Select

Notice Extracted Date

27

March

Accreditation Types

of EA Applications



Entities matched with EAs

11

ANZSICL1 Types



Extracted and Matched EAs

Shareholder(s)

ar Rameshbhai PATEL Keral Kumar Rameshbhai PATEL

ACHHANI, Mayuri | Kamlesh LACHHANI, Mayuri LACHHANI

nz LOPENA | Andresito BOISER, Jason Lorenz LOPENA

BOISER | Andresito BOISER

*** This employer, **MK INVESTMENTS LIMITED - NZBN 9429047109735** is facing the threat of liquidation following an application for liquidation as advertised in the **Gazette on 02 Mar 2026** and filed in the High Court at **Auckland** by **Commissioner of Inland Revenue CIV-2026-404-158**. The company's Directors are **Kamlesh LACHHANI, Mayuri LACHHANI**. The company's Shareholders are **Kamlesh LACHHANI, Mayuri LACHHANI**. Please refer to: <https://gazette.govt.nz/notice/id/2026-aw1070>. Above information raises concerns on the business financial viability under "WA2.10.1 Viable and genuinely operating business". Please carefully consider the above information when processing employer accreditation or any employer-supported work or residence visa application. Confer with employer to get the latest update on the liquidation action and confirm if business is still operating/trading. Request for latest financial accounts - (i.e. Profit & Loss and Balance Sheet statements), along with evidence of debt settlement or debt restructure allowing the company to avoid liquidation proceedings. Check the NZ Companies register (<https://app.companiesoffice.govt.nz/co/7107014>) for latest company registration status. Also, search the name of directors/shareholders if they have been disqualified or prohibited from holding company directorship. A check with the insolvency register (<https://app.insolvency.govt.nz>) and NZ Gazette (<https://gazette.govt.nz>) may also prove useful for any update on the status of the company and for any changes in key persons, both the directors and shareholders. Lastly, consult with your Technical Advisor for the next step action or refer the case to SBFAs for business and financial sustainability checks. ***

*** This employer, **MK INVESTMENTS LIMITED - NZBN 9429047109735** is facing the threat of liquidation following an application for liquidation as advertised in the Gazette on 02 Mar 2026 and filed in the High Court on 02 Mar 2026. The company's Directors are Kamlesh LACHHANI, Mayuri LACHHANI. The company's Shareholders are Kamlesh LACHHANI, Mayuri LACHHANI. Please refer to: <https://gazette.govt.nz/notice/id/2026-404-158>. Above information raises concerns on the business financial viability under "WA2.10.1 Viable and genuinely operating business". Please carefully consider the above information when processing employer accreditation or any employer-supported work or residence visa application. Confer with employer to get the latest update on the liquidation action and confirm if business is still operating/trading. Request for latest financial accounts - (i.e. Profit & Loss and Balance Sheet statements), along with evidence of debt settlement or debt restructure allowing the company to avoid liquidation proceedings. Check the NZ Companies register (<https://app.companiesoffice.govt.nz/co/7107014>) for latest company registration status. Also, search the name of directors/shareholders if they have been disqualified or prohibited from holding company directorship. A check with the insolvency register (<https://app.insolvency.govt.nz>) and NZ Gazette (<https://gazette.govt.nz>) may also prove useful for any update on the status of the company and for any changes in key persons, both the directors and shareholders. Lastly, consult with your Technical Advisor for the next step action or refer the case to SBFAs for business and financial sustainability checks. ***

*** This employer, **STAFF ASSIST LIMITED - NZBN 9429051069605** is under liquidation following the appointment of Insolvency Practitioner as advertised in the Gazette on 02 Mar 2026. The company's Director is **Andresito BOISER, Jason Lorenz LOPENA**. Refer to Document Filings Details <https://app.companiesoffice.govt.nz/co/8573446> Above information raises concerns on the business financial viability under WA2.10.1 Viable and genuinely operating business. Please carefully consider the above information when processing employer accreditation or any employer-supported work or residence visa application. Confer with employer to get the latest update on the liquidation action and confirm if business is still operating/trading. Request for latest financial accounts - (i.e. Profit & Loss and Balance Sheet statements), along with evidence of debt settlement or debt restructure allowing the company to avoid liquidation proceedings, if applicable. Check the NZ Companies register (<https://app.companiesoffice.govt.nz/co/8573446>) for latest company registration status. Also, search the name of directors/shareholders if they have been disqualified or prohibited from holding company directorship. A check with the insolvency register (<https://app.insolvency.govt.nz>) and NZ Gazette (<https://gazette.govt.nz>) may also prove useful for any update on the status of the company and for any changes in key persons, both the directors and shareholders. Lastly, consult with your Technical Advisor for the next step action or refer the case to SBFAs for business and financial sustainability checks. ***

*** This employer, **FIL CONSTRUCT LIMITED - NZBN 9429047444331** is under liquidation following the appointment of Insolvency Practitioner as advertised in the Gazette on 02 Mar 2026. The company's Director is **BOISER**. Refer to Document Filings Details <https://app.companiesoffice.govt.nz/co/7458036> Above information raises concerns on the business financial viability under WA2.10.1 Viable and genuinely operating business. Please carefully consider the above information when processing employer accreditation or any employer-supported work or residence visa application. Confer with employer to get the latest update on the liquidation action and confirm if business is still operating/trading. Request for latest financial accounts - (i.e. Profit & Loss and Balance Sheet statements), along with evidence of debt settlement or debt restructure allowing the company to avoid liquidation proceedings, if applicable. Check the NZ Companies register (<https://app.companiesoffice.govt.nz/co/7458036>) for latest company registration status. Also, search the name of directors/shareholders if they have been disqualified or prohibited from holding company directorship. A check with the insolvency register (<https://app.insolvency.govt.nz>) and NZ Gazette (<https://gazette.govt.nz>) may also prove useful for any update on the status of the company and for any changes in key persons, both the directors and shareholders. Lastly, consult with your Technical Advisor for the next step action or refer the case to SBFAs for business and financial sustainability checks. ***

❑ Additional identifiers exist (e.g., client code or passport number)

Missing client code/passport number

Client Code	First Name	Last Name	DOB	Passport #
null	Emily	Smith	null	null
null	Samuel	Greenwood	null	null

Single individual associated with multiple client codes

Client Code	First Name	Last Name	DOB	Passport #
1234	Grace	Taylor	3/06/1976	AB123456
5678	Grace	Taylor	3/06/1976	AB123456

An individual with multiple passport numbers

Client Code	First Name	Last Name	DOB	Passport #
9101	George	Wilson	19/12/1966	CD123456
9101	George	Wilson	19/12/1966	LY567812
9101	George	Wilson	19/12/1966	RA928375

Director of **XYZ** Ltd



Director of **ABC** Ltd



Same?

Client Code	12345
First Name	John
Last Name	Clerk
Gender	Male
DOB	21/05/1980
Passport #	RN156748

Client Code	67897
First Name	John
Last Name	Clerk
Gender	Male
DOB	?
Passport #	?

❑ Splink is a free and open source Python library for probabilistic record linkage

- Record Linking: link records across different datasets
- Deduplication: identify duplicate records in a single dataset

❑ What Splink can Offer

- Probabilistic matching
- Handles messy, incomplete, or inconsistent data
- Built-in fuzzy comparison functions

id	First Name	Last Name	DOB	Country of Birth
1	Lucs	Anderson	02/01/1984	New Zealand
2	Lucas	Anderson	02/01/1984	Null

Examples of Splink use cases

Use Cases

Here is a list of some of our known users and their use cases:

Public Sector (UK) Public Sector (International) Academia Other

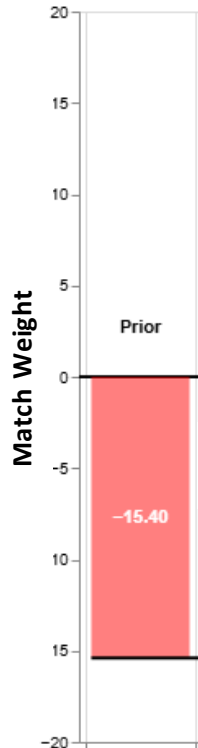
- [Office for National Statistics's Business Index](#) (formerly the Inter Departmental Business Register), [Demographic Index](#) and the [2021 Census](#). See also [this article](#) and [2021 Census to PDS linkage report](#).
- [NHS England](#) is working on developing an alternative data linkage model using splink as the core engine for a new probabilistic data linkage service. This is in order to improve linkage and linkage explainability across NHS datasets. Code now available on [github](#).
- [Ministry of Defence](#) launched their [Veteran's Card system](#) which uses Splink to verify applicants against historic records. This project was shortlisted for the [Civil Service Awards](#)
- [Ministry of Justice](#) created [linked datasets \(combining courts, prisons and probation data\)](#) for use by researchers as part of the [Data First programme](#)
- [Ministry of Justice](#) and the BOLD programme used Splink to power the North Essex Probation Delivery Unit Case Information Dashboard, which won the 2025 Civil Service Award for Excellence in Delivery.
- [UK Health Security Agency](#) used Splink to link HIV testing data to national health records to [evaluate the impact of emergency department opt-out bloodborne virus testing](#).
- The Department for Education uses Splink to match records from certain data providers to existing learners and reduce the volume of clerical work required for corrections
- [SAIL Databank](#), in collaboration with [Secure eResearch Platform \(SeRP\)](#), uses Splink to produce linked cohorts for a wide range of population-level research applications
- [Lewisham Council \(London\)](#) [identified and auto-enrolled over 500 additional eligible families](#) to receive Free School Meals
- [Leicestershire County Council](#) use Splink to match individuals across their Education and Social Care systems. This ensures triage and front-line practitioners have a complete picture of those individuals.
- [Integrated Corporate Services](#) have used Splink to match address data in historical datasets, substantially improving match rates.
- [London Office of Technology and Innovation](#) created a dashboard to help [better measure and reduce rough sleeping](#) across London
- [Competition and Markets Authority](#) identified 'Persons with Significant Control' and [estimated ownership groups](#) across companies
- [Office for Health Improvement and Disparities](#) linked Health and Justice data to [assess the pathways between probation and specialist alcohol and drug treatment](#)

Source: <https://moj-analytical-services.github.io/splink/index.html>

id	First Name	Last Name	DOB	Passport #	Gender	Country of Birth
1	emily	shaw	7/11/1993	NA423891	Female	New Zealand
2	emily	shaw	7/11/1993	RA123587	Female	New Zealand

Match weights waterfall chart

How each comparison contributes to the final score



Overall Accuracy

- Precision (cluster purity): 97.0%
- Recall (client completeness): 99.4%
- F1-Score (harmonic mean of precision and recall): 98.2%

Challenges with NZ Gazette

❖ Data quality

- Name based matching: imperfect in its nature: RegEx may miss new/unusual patterns
- Semi-structured/unstructured source data: Incorrect names, typos, misspellings, etc.
 - **Potential alternative:** *Natural Language Processing (NLP)* can handle messy, inconsistent text, beyond what RegEx can do

❖ Tech and integration

- Python, APIs – internal regulations
 - **Potential alternative:** *Azure Functions* with blob storage and *Azure SQL*

Challenges with Splink

❖ Data quality:

- Missing and inconsistent data

❖ Development:

- Requires thoughtful feature selection and parameter tuning, alternative configurations can lead to variations in outcomes
 - **Potential alternative:** *Python Record Linkage Toolkit/ Dedupe* package

❖ Validation:

- No ground-truth data for validation purposes



Small **operational challenges** can uncover broader **strategic opportunities**



Incremental improvements can lead to **significant transformation** over time



The **quality of data** shapes the value of the insights you can generate



More **connected data** leads to stronger and more actionable risk intelligence



Automation delivers more than **efficiency**



**NEW ZEALAND
IMMIGRATION**

THANK YOU!

datascienceJTA@mbie.govt.nz