MINISTRY OF BUSINESS, INNOVATION & EMPLOYMENT
HĪKINA WHAKATUTUKI

UNCLASSIFIED

Jobs Online – Information about our new data processing pipeline

July 2024

New Zealand Government

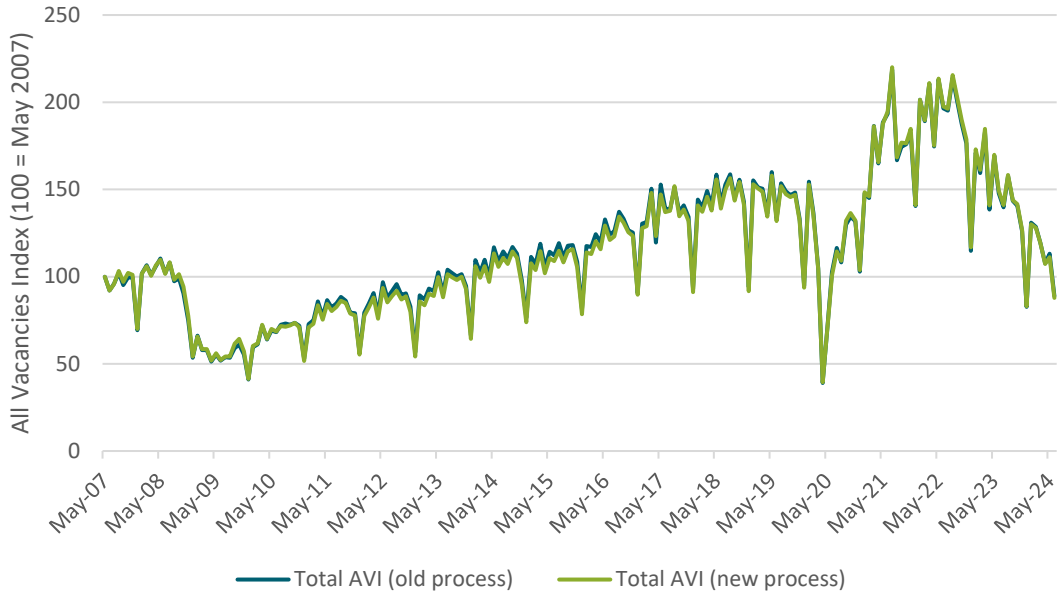
1. Key points

The Ministry of Business, Innovation and Employment (MBIE) is moving to a new data platform. As part of this shift, the Jobs Online data processing pipeline has been redeveloped.

The new process improves the reliability of the data series by incorporating new checks and warnings to better identify any data issues. Efficiency is also improved with the new data processing pipeline requiring less manual input from the person processing the data.

The switch to the new process has not resulted in any major changes to the Jobs Online data series. Figure 1 compares the total vacancies for the new and old processes.

Figure 1: Total All Vacancies Index



Minor changes to note are below:

1. Quarterly data is no longer seasonally adjusted, and so will now show seasonal patterns.
2. The new process no longer removes duplicates across two consecutive months.
3. The new process uses fuzzy matching for mapping job titles to Australian New Zealand Standard Classification of Occupations (ANZSCO). This has resulted in minor changes to the occupation and skill level series.
4. The new data cleaning process and updates to regional and industry coding may result in some minor changes to those series also.

More details on these changes are available in Section 4.

MBIE has done extensive testing and is fully confident in the robustness of all Jobs Online data. However, because of the minor changes noted above, users are advised to re-download the full data series rather than just appending the latest data point.

There will be no interruption to the Jobs Online processing schedule resulting from this processing change.

Link to the latest Jobs Online Quarterly report.

https://www.mbie.govt.nz/business-and-employment/employment-and-skills/labour-market-reports-data- and-analysis/jobs-online/jobs-online-quarterly-report

2. What is Jobs Online?

Jobs Online uses job advertisements as a proxy to measure job vacancies. Jobs Online measures changes in online job advertisements from four internet job boards: SEEK, Trade Me Jobs, Education Gazette and Kiwi Health Jobs. Jobs Online provides a key indicator of labour demand through vacancies by industry, occupation, skill-level, and region.

In December 2009, the first Jobs Online monthly report was released by the Department of Labour, a predecessor of MBIE. The vacancies were sourced from two major job boards – Trade Me Jobs and SEEK. National vacancies were published by five regions, eight industries, and eight occupations. The Jobs Online indices focussed on skilled vacancies (SVI) as internet vacancies from the two job boards were more representative of skilled job ads.

In March 2018, the first Jobs Online quarterly report was released by MBIE. The sources of job vacancies were expanded to also include two specialist job boards – Education Gazette and Kiwi Health Jobs. Moving to quarterly data meant that Jobs Online could now cover both skilled and unskilled vacancies, and allowed a more granular break-down of regional data, supporting in-depth analysis of regional vacancy trends. The Jobs Online indices now focus on the All Vacancies Indices (AVI).

Two versions of the Jobs Online data series are now available:

1. Monthly since May 2007, with breakdowns by five regions, industry, occupation, and skill level.
2. Quarterly since December 2010, at a national and regional level by industry, occupation, and skill level.

3. What was the process for producing Jobs Online?

Jobs Online brings together job advertisements from four online job boards into an index, reported by key breakdowns for the AVI. This section outlines the processes and methods previously used to produce Jobs Online.

1. Data is received

Data files are received within a few days of the end of each month from data providers - SEEK, Trade Me Jobs, Education Gazette and Kiwi Health Jobs. These data files contain all new job ads that were listed in that month.

2. Data is cleaned

A range of manual data cleaning steps are undertaken, including removing blanks and extraneous characters, and ensuring consistent spellings and macron use to help with processing, coding and aggregation of region and industry fields.

3. Duplicates are removed

In some cases, the same job ad might be listed multiple times. Duplicate job ads are identified and removed from the dataset within provider, between providers and across two consecutive months.

4. Vacancies are coded to an occupation standard

Each of the job vacancies is coded to an occupation (4- digit ANZSCO) that enables comparison of the data with other labour market statistics. This was done by an auto-coder which uses a mix of matching techniques that looked for key words in the job title. The results of this occupational auto-coder are regularly audited as part of an iterative process to improve data quality.

5. The Kiwi Health Jobs data is spliced

Kiwi Health Jobs data is only available from May 2011 onwards. To avoid creating a step-change in the published series, a splice is created with healthcare jobs weighted more highly prior to the introduction of Kiwi Health Jobs data.

The original series prior to May 2011 is adjusted by a ratio calculated in the first month Kiwi Health Jobs appeared. The ratio takes account of the difference before and after Kiwi Health Jobs is introduced.

6. Seasonal adjustment of quarterly series is undertaken

A quarterly series was created by summing job vacancies over a three-month period. This quarterly series was then seasonally adjusted to make it easier for users to interpret.

7. Vacancies are presented as an index

Finally, the data is converted to an index rather than job vacancy numbers for publication. An index or percentage change is only published if it includes data from at least two providers.

- The monthly outputs are reported as unadjusted indices with a base month of May 2007.
- The quarterly outputs are presented as seasonally adjusted indices with a base quarter of December 2010.

4. What changes have been made to the Jobs Online process?

The main purpose of the redevelopment of Jobs Online was to improve the reliability and efficiency of the data processing pipeline while minimising any changes to the final output.

As part of this redevelopment, a few methodological changes have been made to the process.

Automation has been increased

Data ingestion, data cleaning and error checking have all been automated. This reduces the manual input and effort required from the person processing the data.

Increased automation has also standardised the accuracy and repeatability of error checking. If errors are detected, the automated process pauses, and the person processing the data can manually intervene as required.

Quarterly series are no longer seasonally adjusted

We removed the seasonal adjustment from the quarterly series and now publish unadjusted indices instead.

The enormous swings in the level of job advertisements due to COVID-19 over the past few years have caused problems for the seasonal adjustment algorithm. After some analysis it was determined the most reliable way of presenting the data would be to switch to using unadjusted indices.

Duplicate job ads between two consecutive months are no longer removed

Duplicate job ads are removed within each data provider's dataset and between different data providers each month.

However, duplicates are no longer removed across two consecutive months as this tended to make the data more volatile by alternately adding and removing any job that was advertised over multiple months.

Occupations are now coded using fuzzy matching

Fuzzy matching was selected as a more scalable solution to replace the auto-coder algorithm previously used for occupation coding to 4-digit ANZSCO codes. This fuzzy matching is multi-layered, using five different passes. The first pass is a "contains" match to a limited number of job titles that have unique ANZSCO codes. The subsequent passes are fuzzy matching to the full mapping file targeting different criteria with each pass.

Before implementing, the fuzzy matching was tested by comparing a matched sample of 200 hand-coded jobs against the old and new occupation coding processes. The fuzzy matching process is of comparable accuracy to the old auto-coder process.

Appendix: Definitions

Advertisement:

Vacancies that are advertised on the job boards of SEEK, Trade Me Jobs, Kiwi Health Jobs and Education Gazette (and other job boards as this data becomes available).

ANZSCO:

Australia New Zealand Standard Classification of Occupations. ANZSCO is a skill-based classification of occupations. The structure of ANZSCO has five hierarchical levels. The highest level (1-digit) contains eight major groups, and each level progressively divides into more specialised occupations.

Fuzzy matching:

An automated program that automatically maps job titles and job descriptions to a 4-digit ANZSCO code using multi-layered fuzzy matching.

AVI:

All Vacancies Index. An index based on all job ads in Jobs Online. The AVI is the main output of Jobs Online and includes all vacancies in occupations that have a skill level of 1-5 under the ANZSCO. This replaced the Skilled Vacancy Index (SVI) as the output for Jobs Online in March 2018.

Data provider:

The data providers for Jobs Online are SEEK, Trade Me Jobs, Kiwi Health Jobs, and Education Gazette (and other providers as this data becomes available).

Duplicates:

Duplicates refer to the duplicate advertisements within data provider and between data provider. Jobs are identified as duplicates if they appear in the same month, have the same region, job title and job description (shortened). Duplicates are no longer removed across two consecutive months.

Industry:

Industry categories are an amalgamation of those used by the respective data providers. The industry is selected from a list of categories by the person placing the job advertisement and is not related to any official industry classification. The industry categories are business, legal, admin and support services; construction and engineering; education and training; health care and medical services; hospitality and tourism; information and technology; manufacturing; transport and logistics; primary industries; sales, retail, marketing and advertising and other.

Jobs Online:

Jobs Online is a package of tools and reports that provide information on the labour market. This includes the All Vacancies Index, a Background and Methodology paper, monthly data releases and quarterly reports on the labour market.

Jobs Online monthly data release:

Data is published monthly, approximately two weeks after the end of the previous month. The data is published as indices with a base month of May 2007.

Jobs Online quarterly report:

The quarterly report uses monthly vacancy data summed over three months. The base quarter is December 2010. The summing enables further in-depth analysis. The quarterly report for Jobs Online is published at the end of every January, April, July and October. This report now uses unadjusted data.

Occupation:

Vacancies that have been advertised on the job boards are coded to 4-digit ANZSCO. The occupation groups reported on are managers; professionals; technicians and trades workers; clerical and administrative workers; community and personal service workers; sales workers; machinery operators and drivers and labourers.

Region:

Five regions: data is available monthly from May 2007 for Auckland Wellington, Canterbury, other South Island and other North Island.
Ten regions: data is made possible by summing over three months, and is available for the following regions from the December 2010 quarter: Auckland, Bay of Plenty, Canterbury, Gisborne/Hawke's Bay, Manawatu-Wanganui/Taranaki, Northland, Waikato, Nelson/Tasman/Marlborough/West Coast, Otago/Southland and Wellington.

Skilled occupation:

ANZSCO has five skill levels underlying its occupational structure. The five skill levels are highly skilled, skilled, semi-skilled, low skilled and unskilled.

Skilled and Unskilled occupations:

An occupation is a skilled occupation if it has an ANZSCO skill level of 3 (a skill level commensurate with an NCEA level 4 qualification) or above. The five skill levels are aggregated into Skilled (highly skilled, skilled, semi-skilled) and Unskilled (low skilled and unskilled).