|  |
| --- |
| **28th Meeting of the Wiesbaden Group on Business RegistersThe Hague, 2 – 6 October 2023** |
| Simon BladesAustralian Bureau of Statistics (ABS) |
| Session 6: New data sources: Opportunity and challenges |
| Down Under Deciphered: Australia’s fresh take on business location  |

# Abstract

In a rapidly evolving global landscape, understanding the geographical dispersion of businesses has become paramount for effective policymaking, supply chain management, and workforce dynamics. The Australian Bureau of Statistics (ABS) has been working assiduously to refine the business location information in the Statistical Business Register (SBR) to meet this need. While the Australian Business Register (ABR) has provided a valuable repository of legal business entities, its primary design as a tax and regulatory tool poses certain challenges for statistical purposes. Notably, the 'head office' effect often presents a skewed representation of actual business locations, particularly for larger entities.

Additionally, the mutable nature of businesses, both in terms of their physical and conceptual locations, demands a more nuanced approach. The ABS has adopted innovative strategies to address these issues, such as implementing a relational database, triangulation from various administrative sources, and a flexible conceptualisation of business location that accounts for emerging trends like remote working. As the ABS forges ahead with these improvements, its experiences offer insight for other organisations grappling with similar challenges in the domain of business location data.

Down Under Deciphered: Australia’s fresh take on business location

Simon Blades – Australian Bureau of Statistics

# Introduction

1. In an interconnected, information-driven world, knowledge of where businesses thrive or decline can be critical in understanding evolving patterns of work, supply chains and the efficacy of policy interventions at a local level. The Australian Bureau of Statistics (ABS) is working to improve business location information maintained in the Statistical Business Register (SBR) to build that critical knowledge base. This paper discusses the ABS SBR in summary, challenges with its business location information, and the strategies the ABS is using now and into the future to improve business location knowledge. This discussion reveals opportunities and challenges that come with trying to understand business location.

# ABS SBR

1. Within Australia, businesses undertaking production are required to register for an Australian Business Number (ABN). ABNs are maintained on the Australian Business Register (ABR) as a register of legal business entities. The ABR is managed by the Australian Taxation Office (ATO) who use the register to monitor actively trading businesses for taxation purposes. The ABR is a rich data source for monitoring business entries and exits, primary business activity (as defined by Australian/New Zealand Standard Industry Classification (ANZSIC)) and business entity information such as activity status. The ABS uses the ABR as the primary data source for its Statistical Business Register (SBR). This is sufficient for most units on the SBR (“non-profiled units”). However, large and complex business structures are not well represented by their ABNs for statistical purposes as the data sourced from the ABR is designed for taxation and regulatory purposes. It is critical to manage a statistical representation of large or complex business structures as the top 3000 businesses in Australia contribute to more than half of wages and revenue reported for taxation. The ABS maintains its own statistical structure of units for these businesses through profiling – the [ABS Economic Units Model](https://www.abs.gov.au/statistics/classifications/standard-economic-sector-classifications-australia-sesca/latest-release#explanatory-notes). Limited information about location is stored for profiled units. Specific addresses are not collected for all locations, but employment in each state or territory of operation is recorded for each industrial unit (Type of Activity Unit).

# Location challenges

1. For non-profiled units the ABR has limitations as an administrative data source, particularly when determining business location. Firstly, business location on the ABR is subject to the ‘head office’ effect for the main business location (a specified address), whereby the only known business location is the mailing address required for taxation correspondence. For larger businesses this could be a head office. The head office may not be in the same location or even state as the main operations of the business. For smaller businesses the address may be an accountant’s office. Additionally, self-reported addresses are subject to quality and consistency errors as until recently, this has been collected as a free-text field rather than registrants selecting from an address register.
2. Furthermore, secondary business locations are self-assigned at registration with the ATO. Until recently, this was limited to a States of operation flag, rather than specified locations, and was collected as a tick box question[[1]](#footnote-1). Consequently, businesses may be optimistic about their future operations and tick some or all states. Alternatively, they may underestimate their future expansion. The States of operation field is not actively maintained by the ATO. This leads to inaccuracies that grow as businesses operate for longer, expanding or contracting from their original States of operations.
3. Prior to the implementation of the ABN by the ATO in 2000, the ABS used profiling to address these location challenges. Replacing this resource intensive method with an administrative data source has had many advantages but came at a cost of relinquishing the ‘Location’ level in the ABS Economics Units Model. At that time, this was an appropriate strategy, as profiling types of activity over many sites and resourcing constraints didn’t have direct utility to main economic indicators at a granular level. More recently, the ABN registration form has been updated so that businesses list their addresses, both main and secondary, using an address lookup. While this is a useful step forward for the future, as newly registered businesses will have much more accurate primary and secondary location information, most businesses on the register have not been updated. Given the poor quality of the primary and secondary location information for businesses in the non-profiled population, it is difficult to use this information for creation of a state level business unit frame.

# Key Strategies for business location improvement

1. With emerging administrative sources creating new opportunities, the ABS is developing options for populating business locations data. Information about activity and employment at business locations has become increasingly important to understand the economic profile of a region. For example, understanding the distribution of businesses, employees and customers during flood or fire events is critical for emergency management. Beyond maintaining high quality profiling of significant units, the ABS has three strategies for improving business location: relational databases, multiple source utilisation and different business location concepts.

## Relational databases

1. As a data structure, the ABS uses a relational database approach to support and organise data beyond the SBR. The relational database is centred around a central list where the ABS SBR is coded with a unique business identifier for each entity within the ABS Economic Units Model. This list is used as a consistent definition of a business unit for statistical processing and provides an important production-defined population for business surveys. However, to ensure the ABS can leverage the power of administratively collected business data, the central SBR list is also linked to other definitions of business used by other organisations. For example, the ATO uses Tax File Numbers (TFNs) for individuals and some businesses to register legal entities for tax reporting. Through linking different definitions of business, the ABS can link business data collected across Australia.
2. After years of building up this relational database with the ABS and using it to inform the SBR there have been several lessons learnt in methodology, data management and SBR innovation. From a methodological point of view there are challenges interpreting data where one-to-many, many-to-one and many-to-many links exist in the data. Prior to a rigorous understanding between lists, the relational database is primarily used as an intelligence and research tool to inform the SBR and experimental administrative data statistical releases. Once the relationship between lists within the relational database is well understood, it becomes standardised when used for statistical processing such as with ABN and the ABS Economic Units Model lists. This approach allows for new lists of information such as geocoding data to be experimented and researched prior to use in official statistics. It should be noted that using many new data sources can increase complication (processing with many variables) and complexity (processing with many connections between variables) of the relational database. Therefore, data management becomes an important consideration. Efforts to describe variables, implement consistent metadata and provide explanatory materials such as process flow diagrams are highly valuable when linking in new data lists. These efforts can increase the initial workload when onboarding data but make downstream processing more far more efficient and more importantly, significantly improves data utility. This is an important investment as downstream processing often is repeated multiple times once data is first introduced into the relational database, compounding the costs of inefficiency in downstream processing.
3. A final critical lesson for building up relational databases is fostering innovation. Statistical products require standards and consistency to be authoritative sources of measurement. When trying to improve on new sources and methods for location information there needs to be space to experiment before products become authoritative sources. The ABS partitions experimental work from statistical production and uses parallel production to provide this safe space to innovate. Experimental releases are regularly made public for feedback from a wide range of stakeholders who can guide further quality assurance, testing and utility. This may mean discontinuing or completely reviewing methods so experimental releases are done quickly to get feedback or only using experimental data as internal intelligence products until methods are more mature. This equally applies to using new information technology infrastructure and big data methods being applied to SBRs. Innovation work is partitioned into a safe space, quickly prototyped, and released for feedback. Lessons learnt from implementing a relational database are not only useful for uplifting business location but can be applied for many statistical processing challenges. It should be a key consideration for national statistical agencies looking to mature SBR intelligence and processing.

## Triangulation

1. The second strategy employed by the ABS is multiple source utilisation where business location is triangulated from different data providers. It builds off the more general strategy of using relational databases to explore business location. While methods for triangulation of business locations is still being matured within the ABS it is hoped that they can help improve quality and timeliness of business location data as well as be powered by big data techniques. In terms of quality, the experimental business locations register is a noteworthy case study that implements data from the ABR, tax submissions and workers compensation insurance registers. The extent to whether these multiple administrative sources align, or deviate, can give greater confidence to the collected address. Data sources may not use ABN or be available for all state jurisdictions which also impacts on confidence and quality. Using multiple sources also breaks the ‘head office’ effect as workers compensation sources require information of where employees are based to offer this type of insurance. The periodicity and scope for these multiple data sources is variable. Datasets of a larger scope are only submitted on an annual basis, whereas smaller collections are updated on a quarterly or biannual basis. This needs to be managed for interpretability, but over time it has the effect of smaller collections supporting more timely disaggregated information and larger datasets supporting overall product completeness.
2. However, business locations work remains experimental. Whilst triangulation is known to increase confidence and fill missing information in a timelier way, further work is needed to quantify the quality impact of these effects. For example, personal addresses used for tax returns may not be up to date, one hypothesis being that young people who are frequently moving house may be more likely to retain a parent’s address or not keep their address up to date with the ATO. Comparing Census residential and employment locations at a point in time is one avenue to investigate. Other data sources such as social security, health department or other addresses may also inform this analysis. As location quality for specific data sources are being evaluated the use of big data to power the strategy of triangulation is also increasingly relevant. In the work of Lenhart et al. (2023) daytime satellite imagery was used to proxy economic activity across periods and spatial units. This novel procedure for determining economic activity would be complementary to administrative collections and profiling activities, providing more timely or nuanced information to the analysis. Furthermore, specific industry activities of businesses such as construction have been identified through overhead Very-High-Resolution imagery (Juergens & Meyer-Heß, 2021). Using different forms of data and analysis from satellite and sensor imagery will require new approaches to maintaining SBRs. There may also be the risk of building ‘black box’ processes as warned by Righi et al. (2019) in their discussion on integrating survey data with big data for official statistics. To implement in the ABS SBR of the future these more advanced forms of integrated techniques require adaptive relational databases and a commitment to triangulation. These big data opportunities may also be enhanced through the interpretation of business location from different viewpoints.

## Business location polysemy

1. The final strategy for improving business location is having an openness to multiple meanings and interpretations of a business’s location. Embracing business location polysemy is achieved by taking a first principles approach to what location information on seemingly unrelated data can say about business location. For example, a conventional view of a business location is the concept of a business property that employees attend to conduct production activities. However, a significant move to working from home in recent years challenges this notion. If workers no longer commute to a central physical space, then the business is now located at their home office according to that conventional view. Policy makers wishing to invest in infrastructure or support specific industries may be more interested in identifying clusters of where employees live rather than their employer location for local investment. This employee clustering analysis may help understand large businesses trading across the country such as supermarket chains. Were this to be linked to data on trade imports, agriculture production and traffic data there is an opportunity to reveal deeper analysis of supply chain dependencies and vulnerabilities.
2. The ABS is looking to develop these alternate location sources and how they fit into a relational database through business location polysemy to realise this potential. Looking at a business at a point in time rather than as a longitudinal entity is another traditional view of business location. As a case study, consider a large telecommunications firm slowly transitioning away from fixed line internet infrastructure to satellite technology over several decades until satellite technology became the primary activity of the firm. The SBR may show the transition as an instantaneous industry movement in the fixed line internet and satellite telecommunication industries that may have implications for technology policy and investment in local areas. The ABS maintains a significant longitudinal register of businesses based on taxation, intellectual property, trade and agricultural sources where researchers can choose to examine business attributes as a point in time variable or smooth for the effect of these SBR changes. This component of the relational database, known as the Business Longitudinal Analytical Data Environment ([BLADE](https://www.abs.gov.au/about/data-services/data-integration/integrated-data/business-longitudinal-analysis-data-environment-blade)) is hosted in a secure environment (the ABS [DataLab](https://www.abs.gov.au/statistics/microdata-tablebuilder/datalab)) with deidentified business units. Academics and government agencies can examine the data using different methods and perspectives for approved research purposes. This provides a trusted user-base for the ABS to trial highly experimental data releases with subject matter experts. By looking at business location from an employee personal address perspective in conjunction with longitudinal business data we can examine a greater scope of sources to understand the world around SBRs.

# Conclusion

1. The pace of information development is an exciting set of opportunities for SBRs. The ABS has been dealing with challenges to collecting business location information such as the ‘head office’ effect, self-reporting and dependence on the ABR. In previous eras, precise business location wasn’t pertinent to the development of statistical products and was not emphasised in profiling activities. Now there are many administrative sources available, with three strategies are being implemented to enrich business location knowledge. First, the SBR being used as a central spine within a relational database to link data across different identifiers and concepts. Second, triangulation techniques are being matured to draw out business location attributes within a SBR based relational database. Finally, the polysemic nature of business location is embraced and not restricted to a ‘bricks and mortar’ concept. The ABS's journey, offers insights for peer institutions globally. As the landscape of business operations evolves, so too must our methodologies, data sources and understanding. The lessons presented herein signal both the progress made and the horizon of opportunities, urging collaborative exploration for a richer global comprehension of business dynamics.

# References

Juergens, C., & Meyer-Heß, F. (2021). Identification of Construction Areas from VHR-Satellite Images for Macroeconomic Forecasts. *Remote Sensing*, 1-12.

Lehnert, P., Niederberger, M., Backes-Gellner, U., & Bettinger, E. (2023). Proxying economic activity with daytime satellite imagery: Filling data gaps across time and space. *PNAS Nexus*, 1-10.

Righi, P., Bianchi, G., Nurra, A., & Rinaldi, M. (2019). INTEGRATION OF SURVEY DATA AND BIG DATA FOR FINITE POPULATION INFERENCE IN OFFICIAL STATISTICS: STATISTICAL CHALLENGES AND PRACTICAL APPLICATIONS. *Statistica & Applicazioni*, 135-158.

1. States of operation is collected in the form of: “Does your business operate at multiple locations? (tick Yes/No). If yes, indicate which states or territories your business operates in (tick states and territories (multiselect))” [↑](#footnote-ref-1)