The Potential for Alternative Data Sources to meet Tourism Data Needs

A report for the Australian Trade and Investment Commission – Austrade and Tourism Australia

Gavin Walker, Christine M O’Keefe, Robert Power, Eleanor McDonald
Report Number EP178217
13 December 2017

Contact: Rod Battye, Tourism Research Australia, Karen Halbert, Tourism Australia
## Contents

Acknowledgments ........................................................................................................................... 3

About the authors ......................................................................................................................... 4

Executive summary ....................................................................................................................... 5

1 Introduction ........................................................................................................................ 9
   1.1 Background..................................................................................................................... 9
   1.2 Note on the scope of the project .............................................................................. 10
   1.3 Structure of this report ............................................................................................. 10

2 Findings from tourism data consumer consultations .............................................................. 12
   2.1 Users of tourism data, and broad purposes – tourism data consumer view .......... 12
   2.2 Privacy ......................................................................................................................... 15
   2.3 Review of other countries’ tourism data practices .................................................... 16
   2.4 Assessment of current methods ............................................................................... 19
   2.5 Sentiment of tourism data consumers .................................................................... 23

3 Alternative data sources ...................................................................................................... 25
   3.1 Overview of data use in other industries ............................................................... 25
   3.2 Alternative data source classes ................................................................................ 28
   3.3 Alternative data sources to inform tourism statistics ............................................ 31
   3.4 Risks of using ADS ..................................................................................................... 44

4 Roadmap for inclusion of alternative data sources ............................................................... 46
   4.1 ADS based services .................................................................................................... 46
   4.2 Funding ADS services ............................................................................................... 49
   4.3 Changing the survey operation ................................................................................ 50

5 Conclusion ............................................................................................................................. 52

Glossary 53

Abbreviations 54

References 55
Figures

Figure 1 Tourism data roadmap.............................................................................................................................................. 7
Figure 2 Exponential growth of data generation...................................................................................................................... 26
Figure 3 Current employment data ecosystem. ....................................................................................................................... 28
Figure 4 Proportions of number of transactions by transaction type (Doyle, et al 2017)......................................................... 36
Figure 5 Twitter usage in Australia based on a sample of people aged 14+ (Roy Morgan 2015).................................................. 38
Figure 6 Percentage of Australians who used a travel agent in 2016 (Roy Morgan 2017)......................................................... 40
Figure 7 GDS air market share (Business Travel IQ 2017)........................................................................................................... 41
Figure 8 Tourism data roadmap.............................................................................................................................................. 48

Tables

Table 1 Qualitative assessment of utility of classes of alternative data sources for tourism data types................................................. 6
Table 2 Utility of tourism data source classes in each of the types of tourism data ................................................................. 32
Table 3 Percentage (%) of trip cost prepaid by country of origin and purpose 16/17 (TRA) ............................................................ 34
Table 4 Payment timing for overnight domestic trips 16/17 (TRA) .............................................................................................. 34
Table 5 Market share of credit card providers by size of card balances (APRA 2017) ................................................................. 36
Table 6 Count of active Australian Facebook users (Cowling 2017) .......................................................................................... 39
Table 7 Mobile telecommunications market share and coverage (Productivity Commission 2017a) ......................................................... 43
Acknowledgments

This project was funded by Austrade and Tourism Australia. The project team was assisted by Tourism Research Australia.

State and Territory Tourism Organisations, as well as other tourism data stakeholders and tourism data subject matter experts, provided valuable information and insights during the consultation phase of the project. We thank them for generously making time available to speak with us, and for their patience in reviewing and correcting our notes taken during the consultations. This report is richer with their inputs.

The authors thank their colleagues James McHugh and Bella Robinson for guidance and advice on how social media can be used for tourism purposes and for details of the CSIRO social media monitoring tools. Further, the authors acknowledge contributions from Simon Barry, Simon Dunstall and Andrew Reeson of CSIRO Data61.
About the authors

Gavin Walker is a Principal Experimental Scientist at CSIRO Data61. Gavin has over 25 years of experience in distributed information systems research and development. He has worked in a broad set of domains including: spatial information systems, environmental informatics, emergency services, traffic simulation, sensor clouds and museums. Through the Water Information Research Development Alliance between CSIRO and the Bureau of Meteorology he developed the Water Data Transfer Standard and made significant contributions to the WaterML2.0 Open Geospatial Consortium Standard. Gavin was a technical contributor and an author of the Productivity Commission draft report on Data Availability and Use.

Dr Christine O’Keefe is a Senior Principal Research Scientist in CSIRO Data61, and an Adjunct Professor in Mathematics at the University of Adelaide. Since 2004, Christine has pursued research in the development and evaluation of methods to address the balance between allowing data access and use with protecting the confidentiality of people and organisations represented in data. Christine has contributed to the design of remote access systems as well as measures of disclosure risk, information loss and data utility. As well as making theoretical contributions, Christine has worked on practical implementations while seconded to the Australian Bureau of Statistics in 2011-2012, and has provided advice to various government agencies including the Office of the Australian Information Commissioner, the Australian Government Department of Health, and the Australian Government Department of Climate Change and Energy Efficiency.

Robert Power is a software engineer and project manager who has been with CSIRO since 1992. His research and applications interests focus on information management, data integration, spatial databases, web systems and natural language processing applied to social media content for situational awareness. He has worked in numerous domains, ranging from astronomy, privacy and security, water informatics, government services and emergency management.

Dr Eleanor McDonald has a PhD in Numerical Analysis from the University of Oxford and, prior to this, worked at an environmental engineering consultancy specialising in acoustic and air quality modelling. Since joining Data61 in 2016, she has worked on a variety of projects in the area of statistical modelling and monitoring as well as text analytics. She has worked with several government agencies including the Australian Bureau of Statistics and Australian Security and Investments Commission and has also worked on several projects examining bushfire risk, both in Australia and overseas.
Executive summary

The Australian Trade and Investment Commission – Austrade – seeks to advance Australia’s tourism interests by providing policy, information, advice, and services. Austrade is responsible for conducting Australia’s International Visitor Survey (IVS) and National Visitor Survey (NVS), the primary instruments for monitoring inbound and domestic tourism activity and trends for the purposes of policy and program development, business planning, and tourism marketing by all levels of government and the private sector.

Austrade, through Tourism Research Australia (TRA), focuses on continually improving its tourism data offering, seeking to explore beyond its traditional sources and considering the full range of data sources available to the tourism sector. In particular, with an increasing amount of new information becoming available through administrative and digital data sources, Austrade has asked Data61 to carry out a high-level and exploratory investigation into the data requirements of tourism data consumers, how well mainstream data products (including IVS, NVS and satellite accounts, see section 2.4) meet those needs and the suitability of Alternative Data Sources (ADS) for use in tourism statistics. Data61 will then articulate a roadmap for the provision of tourism data.

Data61 carried out interviews with all State tourism offices and with a range of tourism industry representative bodies (see section 2.1.1) to ascertain their data requirements, how well mainstream tourism data products met those requirements and confirmed or proposed engagement with ADS.

Their tourism data requirements were categorised into five types (Accommodation, Spend, Movement, Purpose and Impact, see section 2.1.2) and ADS categorised into four classes (Transactions, Social Media, Booking/Distribution/Ticketing and Telecommunications, see section 3.2). A major data provider from each ADS class was investigated to understand that ADS class. Each ADS class was qualitatively assessed against its utility to provide statistics for the five tourism data types based on a range of quality criteria including: Institutional Environment, Accuracy, Latency, Coherence, Interpretability, Scalability, Linkability, Coverage and Representativeness (see section 2.4.1).

Key findings

1. **Mainstream data products.** Existing mainstream data products are generally meeting the needs of tourism data consumers for the purposes (reporting, supporting business, marketing and tourism activity) and the spatial (national, state and some regional areas) and temporal (rolling annual, quarterly releases) scales for which they were designed. However the data products could be made more accessible to those who are not tourism data specialists. Some gaps include: understanding visitor pathways, understanding visitor sentiment including experiential expectation and actual experience, and understanding the cohort of people who have not travelled, particularly what could help them decide to make a trip.

2. **Expectations increased.** Consultations with tourism data consumers have indicated expectations of what types of data should be available have increased compared to what they receive from mainstream data products. Some tourism data consumers are seeking data,
especially for marketing purposes, at finer spatial and temporal scales. The mainstream data products are not meeting these needs. There appears to be lack of clarity about the utility of mainstream products at these spatial and temporal scales. This can lead to the inappropriate use of mainstream products.

3. **ADS offer potential.** ADS offer potential to provide data at finer spatial and temporal scales for many tourism purposes. In practice privacy and commercial concerns mean that current practical examples of ADS-based data products operate over a small pool of data. For example, NAB represents only 14% of the credit card market and may not be willing to join their dataset with Westpac (a 23% share) and the Commonwealth Bank (a 27% share) to achieve better coverage. These restricted ADS services are unlikely to perform well across a range of quality criteria but especially in Coverage, Coherence, Representativeness and Linkability (see section 3.3.1). Data61’s consultation with tourism data consumers suggested they are yet to understand the quality issues related to the results that may be produced by an ADS.

The table below summarises the utility of a class of ADS. It is based on a qualitative assessment of individual ADS classes against the quality criteria described above. A high rating suggests that the ADS class is likely to be useful in providing tourism statistics for the tourism data type. A medium rating suggests some utility might be obtained. A low rating suggest the class is unlikely to provide useful information. No rating means the data class does not provide useful data. An ADS service that could link data across individual ADS classes could combine their utility. The existing mainstream tourism data products are high quality in all tourism data types. In this qualitative assessment Data61 is not directly comparing the ADS against the mainstream products. That is best done in the context of the purpose and scale of the data use and when both sets of data are available for comparison.

<table>
<thead>
<tr>
<th></th>
<th>Accommodation</th>
<th>Spend</th>
<th>Movement</th>
<th>Purpose</th>
<th>Impact</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Transactions</strong></td>
<td>Low</td>
<td>High</td>
<td>Medium</td>
<td>Low</td>
<td></td>
</tr>
<tr>
<td><strong>Social Media</strong></td>
<td></td>
<td></td>
<td>Low</td>
<td>Medium</td>
<td></td>
</tr>
<tr>
<td><strong>Booking/Distribution/Ticketing</strong></td>
<td>High</td>
<td>Low</td>
<td>Medium</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Telecommunications</strong></td>
<td>Low</td>
<td></td>
<td>Medium</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

4. **5-10 year time series.** Most tourism data consumers do not want to see a sharp change in tourism data where there is a significant change in quality and a 5-10 year time series is not available. Data is fundamental for making decisions, the quality of the data needs to be understood.

**Summary of key recommendations**

1. **Communicating uncertainty.** Communication and education around the uncertainty of the mainstream products for common purposes and the impact of the choice of spatial and
temporal scales needs to be improved. Materials need to be accessible to CEOs and tourism groups, not just state tourism researchers.

2. **Deploy ADS services.** In a series of workshops over the first year tourism data consumers, ADS custodians and ADS infrastructure suppliers should workshop practical deployments of ADS services that maximise the linkages (and therefore the richness of the data pool) between ADS. The ADS services should ideally contain data from multiple sources within an ADS class and sources across ADS classes that are linked on individuals. For example, a combined Westpac (transaction), NAB (transaction) and Telstra (telecommunication) ADS service. These ADS deployments would be clearly named, to differentiate them from mainstream data. These services would allow access to aggregated, deidentified and confidentialised data through a wide set of operations to an agreed set of tourism data consumers. This would require seeking significant involvement from the private sector. There is less utility in deploying an ADS service with a single ADS class (such as transactions) and a limited set of operations (such as spend/month/region). In the workshops the needs of different tourism data stakeholders will need to be prioritised. All deployed services should use an equivalent definition of tourism. It may take until the end of the third year to complete the deployment of the services agreed in the first year.

3. **Explore ADS funding.** Explore options to fund the deployment of ADS services and their evaluation. This can happen in parallel with the ADS deployment workshops. Cutting existing mainstream tourism data services over the next five years would be detrimental to the evaluation of ADS. Importantly, it would also deprive the tourism community of reliable, quality data, with a time series, while the utility of the ADS services are being determined. The results of the five year review may indicate that it will be important to further continue mainstream services if the ADS services require additional evaluation, for example, if the initial ADS services are changed significantly.

4. **Evaluate ADS services.** Evaluate emerging ADS services against the mainstream data products to identify their strengths, weaknesses, utility and niche. Services with wide access by tourism data consumers will garner the best evaluation over an extended period and assist the tourism data consumer community in understanding the role and utility of ADS. Experiences would be workshopped annually and inform training material for the ADS. The training material for ADS would be of a similar nature to the mainstream training material, highlighting how and where the data services are best used. The ADS services would be reviewed in the fifth year from the start of the initial workshop.

![Figure 1 Tourism data roadmap](image-url)

The Potential for Alternative Data Sources to meet Tourism Data Needs | 7
1 Introduction

1.1 Background

The Australian Trade and Investment Commission – Austrade – seeks to advance Australia’s tourism interests by providing information, advice, and services. Austrade achieves this by generating market information and insight, promoting Australian capabilities, developing policy, making connections through an extensive global network of contacts, and providing quality advice and services. Specifically, Austrade provides advice to the Australian Government on its trade, tourism, international education and training, and investment policy agenda, and develops policy, manages programs and provides research to strengthen Australia’s tourism industry and to grow Australia’s tourism market share.

Tourism is often seen as a demand side concept, with the tourism industry being defined by the products visitors purchase before (e.g. pre-paid travel packages, cameras, batteries, memory sticks, and suitcases) during (e.g. cash or credit card purchases) and after their travel (e.g. duty free products in their home airport) and the locations they visit.

The tourism industry is therefore a complex amalgamation of contributions from a range of industries supplying numerous products and services (e.g. air travel, food, accommodation, fuel, taxi, bus and rail transport, entertainment, alcohol and beverages, telephone services) consumed by visitors associated with their trip. Some of these expenditures are paid by visitors themselves and some may be paid on behalf of visitors by someone else such as a friend or relative, government, or local body. Tourism Research Australia (TRA), a branch within the Tourism Division of Austrade, is Australia’s leading provider of quality tourism intelligence across both international and domestic markets. This status is underpinned by its robust tourism data collections and broad ranging research and analysis program.

Austrade is responsible for conducting Australia’s International Visitor Survey (IVS) and National Visitor Survey (NVS). These surveys are the primary instruments for monitoring inbound and domestic tourism activity and trends for the purposes of:

- Reporting and tracking tourism performance
- Policy and program development
- Business planning
- Tourism marketing

Key demand side metrics of tourism include visitor numbers, visitor nights and visitor expenditure. The overall economic value of the tourism industry activity is measured by collating detailed expenditure (spend) information from TRA’s IVS and NVS surveys and other sources and calibrating these with the industries supplying these products via the national accounting framework. The Australian Bureau of Statistics (ABS) Tourism Satellite Account provides estimates of tourism’s contribution to Australia’s Gross Domestic Product (GDP), Gross Value Add (GVA) and employment.
Tourism supply is measured by integrating information from monitoring of accommodation (room supply), the ABS tourism satellite account (for direct and indirect labour force supply), and BITRE (Bureau of Infrastructure, Transport and Regional Economics) data (for domestic and international aviation capacity).

Austrade, through TRA, is focused on continually improving its tourism data offering. Given this, it is seeking to explore beyond traditional sources by considering the full range of data sources available to the tourism sector. In particular, with an increasing amount of information becoming available through administrative and digital data sources, the question arises whether Alternative Data Sources (ADS) (section 3) could satisfy tourism data consumer needs more effectively than is currently achieved by mainstream tourism data.

Currently, many governments, industries and businesses interested in tourism are investigating the potential of ADS including data from mobile phones, credit cards, reservation systems and social media platforms, to supplement or even replace mainstream data sources.

1.2 Note on the scope of the project

In its role as trusted advisor to Government, CSIRO Data61 brings its scientific integrity, broad experience in data and expertise in seeking the best interests of Australia. Data61 was invited by Austrade and Tourism Australia to conduct a high-level and exploratory investigation into:

1. The data requirements of tourism data consumers.
2. How well mainstream data products (including IVS, NVS and satellite accounts) meet those requirements.
3. Overseas experience with tourism data and ADS.
4. The suitability of ADS for use in tourism statistics.

Then from these develop a roadmap for the future role of ADS in tourism. Tourism data consumer requirements and satisfaction were ascertained though interviews, see section 2 for a detailed description of the methodology. The ADS assessment involved classifying ADS data sources into classes and contacting a significant data provider in each class or examining descriptions of their data offerings on public web sites. See section 3.3 for a detailed description of the methodology.

1.3 Structure of this report

The remainder of the report is structured as follows.

Section 2 summarises the information we gathered during the tourism data consumer consultations, with an emphasis on the contributions that we found most useful for the project. It also reports the findings of related approaches in a small selection of other countries and other industry sectors. Section 2 concludes with an assessment of the capacity of current tourism data collections to meet the data needs of tourism data consumers.

Section 3 provides a discussion of the potential suitability and benefits of alternative data sources to better meet tourism data needs, including an outline of the risks associated with using alternative data sources, and suggestions of risk mitigation strategies. Informed by the foregoing
analysis, a roadmap for exploring a move to alternative data sources is provided in section 4. A conclusion is presented in section 5.
2 Findings from tourism data consumer consultations

This section summarises the information gathered during the tourism data consumer consultations. It draws together consumer data requirements into a set of five tourism data types which are used in section 3 to assess the utility of ADS. This section also reports approaches in a small selection of other countries. The original notes from the majority of individual consultations, corrected and edited by the tourism data consumer representative involved, have been provided to Austrade and Tourism Australia for reference.

The section concludes with an assessment of the capacity of current tourism data collections to meet the needs of tourism data consumers. It does this by introducing quality criteria which are used to formalise the assessment of mainstream data sources by tourism data consumers. In section 3 the same quality criteria are used to assess the ADS.

The IVS and NVS were created 25-30 years ago to enable the production of national accounts quantifying the contribution of the tourism industry to the Australian economy. Surveys are widely acknowledged to be the best way to provide consistent, robust and reliable indicators of tourism activity, as are required for the production of the Australian Tourism Satellite Account by the Australian Bureau of Statistics. These statistics have been used for business (and infrastructure) planning, and tourism marketing. Surveys capture a full picture of the tourism experience as all the facets of a tourism activity are tied to a single tourist.

As breadth and speed of data capture and use has increased over the last twenty years, tourism data consumers have become accustomed to the specificity and immediacy of information. They are seeking to use the source of trusted and accurate tourism information (the surveys and satellite accounts) to pursue these purposes in smaller geographic regions and to measure rapidly changing events close to when they unfold.

Naturally, this has led to dissatisfaction amongst some tourism data users who are unsurprisingly finding that the survey outputs are not well-suited to these fine spatial and temporal scales, and this has been seen in our consultations. Understandably, such data users are starting to look for solutions to this challenge, including advocating, considering or actually moving to explore the potential of ADS.

2.1 Users of tourism data, and broad purposes – tourism data consumer view

This section lists the tourism data consumers consulted, then describes their view of what they use the data for, and what they do with information. Most of the findings are of a general nature as some of the tourism data consumers chose not to be named in the report.
2.1.1 Tourism data consumers consulted

Each consultation took place over the phone or in person over the course of an hour. A framework of questions was used as a broad guide to the conversation. Interviewees were asked to discuss their data requirements, how well mainstream data products met their needs and their engagement with ADS. The latter was primarily to determine ADS of interest to tourism data consumers.

The set of tourism data consumers was drafted from initial consultations with TRA and TA. TRA and Data61 further refined the list. It was agreed between TRA and Data61 that the consultations would cover State Tourism Offices, Tourism Industry Bodies, ORC International (who designed the IVS and NVS) and Statistics Canada. Tourism Industry Bodies (such as Tourism Accommodation Association) were chosen over specific industry data consumers because they could give a broad view of the requirements of the industry. Picking specific industry participants may have been unrepresentative of the broad view of that industry. The exception was the Mantra group whose CEO was interviewed at the request of TRA.

Between June and September 2017, Data61 consulted with the following:

Category 1: Commonwealth Tourism Organisations
– Tourism Research Australia, Tourism Australia

Category 2: State and Territory Tourism Organisations (STOs)
– Tourism NT, VisitCanberra, SA Tourism Commission, Tourism and Events Queensland, Destination NSW, Tourism WA, Tourism Tasmania, Tourism Events and Visitor Economy (Victoria)

Category 3: Tourism business operators (or representing bodies)
– Tourism Accommodation Australia, Restaurants and Catering Australia, Tourism Chambers and Restaurant and Catering Australia, Association of Australian Convention Bureaux, Australian Tourism Export Council, Australian Hotels Association of WA, Mantra Group, Australian Regional Tourism Network

Category 4: Others
– ORC International, Statistics Canada

Each tourism data consumer’s view was influenced by their category. For example, those in State Government had purview of the regions in the state. In addition, each tourism data consumer was keen to have more information relevant to their own sector and objectives.

2.1.2 Tourism data consumers’ purposes for using tourism data, and data demands

All tourism data consumers acknowledged the usefulness of the IVS and NVS, and the current lack of any viable alternative. The most common data needs not met by the surveys were: lack of robust and reliable estimates at regional levels (for example, visitor motivation to travel to a regional tourism destination), lack of estimates at finer temporal scales (such as monthly or weekly) and length of time between data collection and release of the data products. Some tourism data consumers pointed out gaps in the IVS and NVS coverage, such as, the cruise
sector and international travellers using airline member lounges. TRA reports to have addressed many of these gaps, however the perception of gaps remains. Others wanted to understand more about tourist motivations to better tailor products, especially in regional areas.

The tourism data consumers identified a broad range of uses for tourism data, including:

– **Reporting:** Meeting government reporting requirements; evaluating performance against Tourism 2020 targets; ministerial briefing; comparison of performance of the tourism sector across states and/or regions; quantification of indirect benefits of tourism (e.g., cherries picked by backpackers on casual employment)

– **Business support:** Supporting the business needs of members including problem solving; investment planning for infrastructure, promotion, and facilities; tourism industry supply capacity planning; tracking tourism-related performance of events

– **Marketing:** Targeted marketing, including development of typical tourist `personae’

– **Tourism activity:** Understanding the behaviour of tourists, including: where they went (including dispersion), how long they stayed, what they did, what they spent, what they want to do; understanding tourist `pathways’; measuring return on investment by public and private sector and supporting lobbying for investment; understanding the cohort of people who have not travelled (why not, what could make them decide to travel)

Most of the tourism data consumers expressing these views were interested in business planning and marketing purposes, at very fine spatial and temporal scales. Daily data and at SA2 (roughly a suburb in a city) spatial scale were said to be desirable, but realistically monthly and Local Government Area (LGA) would be enough for most. When questioned further consumers acknowledged that the surveys were designed for giving a national to state picture at a quarterly interval, rather than for these much finer scales.

In a recent paper to TRA on tourism indicators, ABS suggested an annual indicator suite, with some values available quarterly would give the best quality and cost trade off. These products are high quality but perhaps there are other avenues for finer spatial temporal scales. Many of the tourism data consumers believe there is a potential for ADS to fill the gaps they identified.

To meet these purposes, the various tourism data consumers make use of data including:

- **Accommodation**
  - Accommodation supply, use, performance and profitability across different accommodation products

- **Visitor expenditure**
  - broken down by origin market, purposes or segments (e.g. holiday, friends, conventions, incentives, food and wine) and expenditure class (e.g. accommodation, attractions, food, beverage)

- **Movement**
  - Transport mode, especially airline seats (capacity) and volumes
  - Visitor numbers: international, interstate and intrastate with destination, origin and home location.
  - Numbers at attractions (e.g. events, theme parks)
• Purpose for visit
  – Sentiment and experiential expectation (what sort of experience are they after in terms of effort and enrichment)
  – Demographics
  – Booking source

• Sector impact
  – Tourism related employment models
  – Business entries and exits at postcode level
  – Industry sentiment

These data items are found in the IVS and NVS surveys and the satellite accounts. However there is more demand for greater representation at finer scales. Some other data items, including airline passenger movements and Overseas Arrivals and Departures (OAD) data, and some accommodation data, are routinely collected from the relevant administrative systems. They are integrated into the IVS and the satellite accounts.

The gaps most commonly mentioned included: understanding visitor pathways, understanding visitor sentiment including experiential expectation and actual experience, and understanding the cohort of people who have not travelled, particularly what could help them decide to make a trip.

There was some evidence of a disconnect between the information products generated by TRA and the way this information is interpreted and used by some tourism data consumers. This largely reflects issues around the understanding and interpretation of the data’s more technical aspects.

Most tourism data consumers were interested in the potential of ADS to fill the gaps between their data needs/desires and what is currently available to them. For many tourism data consumers, the most pressing issues were spatial and temporal detail, latency, and accuracy of data. These are discussed in more detail in section 2.4.

2.2 Privacy

The value of tourism data can only be realised if it is able to be shared with those who have the necessary expertise to analyse and use it. Since tourists are people, tourism data contains personal information, and so sharing or releasing it has both legal and ethical implications, as well as being subject to community expectations, regarding privacy protection. Tourism data consumers demonstrated a good understanding of privacy issues, including:

• Whether tourists are happy that their data is being used for other purposes? Do they know?
• If data users are unable to access data suppliers’ data directly, how can they be sure the data products are correct, and are fit for purpose?
• How could we gather useful information, for example, with names and addresses for linkability, and still protect tourists’ privacy?
A robust and thorough Privacy Impact Assessment (PIA) regime¹ should accompany any tourism data use, from proposal to completion. Informed by such a PIA, a data management strategy needs to be developed, including governance, infrastructure, and privacy risk assessment and management. An initial issue to resolve would be who should take the lead in addressing privacy concerns in tourism data use. Maintaining the trust of the public is essential to ensure robust data supply.

A recent framework provides a practical guide to de-identification and decisions that need to be made in using it as a tool for privacy protection (O’Keefe, et al 2017).

2.3 Review of other countries’ tourism data practices

2.3.1 New Zealand – Tourism Data Domain Plan and Regional Tourism Indicators Review

The New Zealand Tourism Data Domain Plan (NZ 2011) was prepared in 2011 by the Tourism Data Review Project Team at the Ministry of Economic Development to guide the compilation, dissemination and use of tourism statistics. It was developed in consultation with tourism data consumers, including tourism businesses, government agencies and tourism industry bodies. The plan was designed to achieve clarity and agreement from tourism data consumers about the main priorities for tourism statistics, and provide the strategy for addressing the priorities over the next five to eight years.

The plan contemplated the potential to use electronic card data (credit card, EFTPOS) to address issues in tourism-related research. Such data was scored against its potential usefulness in informing each of five topic areas, as follows (with 1 being most useful and 7 being least useful). Electronic card data was scored at 2 for Value of government investments, 4 for Value of tourism to NZ, 5 for both Global competitiveness, and Growth, innovation, productivity and efficiency of tourism businesses, and 7 for Sustainability of tourism.

Interestingly, the plan recommended the development of regional indicators of tourism from alternative data sources (Recommendation 2). In response, the (New Zealand) Regional Tourism Indicators (RTIs) were developed in 2012 to measure the value of tourism in the regions. The name ‘Indicator’ reflects the fact that the RTIs do not actually estimate a spending figure for several reasons, including: monthly measures may have more variations that do not follow long term trends and so the explicit indexing back to the average month in 2008 helps centre user interpretations of small movements; raw RTI spend represents only a sample of electronic spend which is difficult to conceptualise; and other data to mitigate these problems is only produced annually.

The New Zealand Regional Tourism Indicators Review – Final Recommendations Paper (NZ 2015) was released in 2015. The review recommended ceasing to publish indexes and starting to publish monthly provisional estimates, in response to users’ need for spend estimates, and level of comprehension of how to interpret the RTIs, despite potential risks such as the estimates being

noisy yet used as the ‘truth’, causing subsequent discomfort when they were updated or different from other estimates (Recommendation 1).

Recommendation 5 in the same paper focussed on improving management of ‘Paymark’ merchant changes affecting the international RTI series. The issue was that purchases made by international credit cards at a merchant using the Paymark network were included in the International RTI series, however, the number of merchants using the Paymark network was constantly changing. Standardised systems and rules would be needed to manage the data to ensure there was no adverse effect on the validity and reliability of the international RTI series. It is also necessary to cover the gap between the sample frame (about 70% of merchants) and full coverage by imputing a non-Paymark spend at an agreed sub-national level, which will mitigate the effect of disruptively large changes in market share.

Other recommendations included to develop datasets that look at visitor satisfaction and decision-making processes, and ensure that the data is accessible and aligned to other initiatives (Recommendation 8), and to investigate the use of social media and user-generated media (e.g. TripAdvisor) to monitor and provide feedback on global comparisons and competitiveness (Recommendation 9).

In summary, NZ have been able to trial their ADS based tourism data in a way that allowed the tourism sector to explore the utility of that source. They found significant variability in the coverage of the data made it hard to provide reliable data on a monthly basis. However, this did not reduce the demand for even more data. It highlighted the need for education, communication and transparency in providing data.

2.3.2 Canada – Interview with Statistics Canada

Tourism data in Canada comprises administrative data (border counts) plus data from the National Travel Survey (NTS) and International Travel Survey (ITS).

For the NTS, information on domestic and outbound travel is collected from Canadian residents. Given the objective of cost-efficient, effective collection of quality data, Canada has piloted a web questionnaire to increase self-response with minimal interviewer intervention. They are working with the Canada Border Services Agency to understand how to confidently utilise contact details of Canadians travelling abroad for the purpose of better targeting the sample.

The ITS (redesigned into the Visitors Travel Survey (VTS)) collects information from international visitors to Canada. However some challenges have arisen because of the lack of a survey framework for interception at the border that has agreement with Border Services.

For both surveys, Canada is aiming to collect data directly at tourist sites across the country. Canada is also trying to pilot more structured/targeted approaches and raising response rates. Consideration is being given to use incentive schemes to encourage respondents to participate in the survey.

In addition, as part of Statistics Canada’s modernisation agenda, the aim is to rely more and more on administrative and alternative data sources, rather than surveys. Statistics Canada is actively looking at new ADS, such as financial transaction data, and conversations with ADS holders to obtain aggregate tourism information are ongoing. They are starting to explore more formally
whether social media data can be used to measure traveller activity. However, based on preliminary discussion with Artificial Intelligence experts, it’s not clear how they could obtain information on tourism expenditures. Initial conversations suggest social media data may be more helpful to give information on proportions and ratios relative to a reference population or baseline that is obtained in another way. This could be suitable for indicating trends in terms of proportions if data users were comfortable with overall estimates. Further exploration and clarification from experts are acknowledged by Statistics Canada to be required.

In summary, Statistics Canada is exploring the use of administrative and alternative data sources. They plan to continue investigating how ADS can be used to enrich survey data, and possibly to reduce respondent burden by, for example, reducing the length of the interview. Challenges include: classifying transactions as related to tourism or not, recall bias, declining survey response rates, and integration of multiple data sources from survey and alternative data sources.

### 2.3.3 Spain – Analysis of linked telecommunications and banking data

In 2014 in Spain, Telefónica and RocaSalvatella conducted a study (Sust 2014) designed to be a first step in understanding the possibility of big data to contribute and add value to the Spanish tourism sector. In particular, they investigated the potential to incorporate data collected from the electronic activity (financial transactions and phone use) of anonymous foreign tourists into market research for the hotel industry in particular, thereby providing a new methodology for improved analysis and knowledge of the Spanish tourism industry.

The study brought together data about the activities of foreign handsets using Telefónica’s infrastructure between 7 and 21 October 2012 in Madrid and Barcelona; with data on electronic payments by foreign cards sent through the BBVA bank’s terminal network in those cities during the same period. In total, 680,928 mobile handsets and 168,921 cards were included in the study, associated with visitors from 21 different countries. Analysis of anonymised, aggregated data underpinned the report’s conclusions and recommendations.

Although absolute numbers could not be obtained from the data, the study demonstrated that analysis of the data could provide information to assist with:

- Winning more customers by pinpointing the countries, channels and languages on which it is recommended to focus marketing
- Detecting areas of the city in which commercial transactions are carried out, e.g. accommodation
- Developing an attractive product suited to customers’ true needs (ideal length of package offers, information about complementary services demanded by different nationalities, etc.)

On the other hand, it was also acknowledged that any decision in this area should be based on more in-depth analysis covering a longer timeframe, which would make it possible to show seasonal and annual variations. The study showed real insight could be gained from the use of ADS. A great strength of the study was that it linked a strong movement source (phone records) and a financial source (electronic transactions). Such linkages can significantly increase the value and accuracy of ADS.
2.4 Assessment of current methods

The following are considered to be the mainstream data sources:

- International Visitors Survey
- National Visitors Survey
- Overseas Arrivals and Departures
- International Visitor Cards
- Migration
- Flight Schedules and Capacity
- Australian National Public Toilet Map
- Regional Airport Operations
- Labour Force Survey
- Australian Industry Statistics (cafes, clubs, pubs, accommodation, fuel retailers and transport)
- Household Expenditure Survey

These sources include both survey and administrative data. They are used to create the TRA data products and the ABS Tourism Satellite Accounts.

While tourism data consumers were asked about their data needs in relation to the IVS and NVS in particular, Data61 holds that the concerns expressed are transferable to other data sources as well.

2.4.1 Quality framework

This project is examining the utility of existing tourism products as well as ADS. The quality of the data sources is an indication of their utility. The ABS data quality framework (ABS 2009) defines seven criteria for indicators: institutional environment, relevance, timeliness, accuracy, coherence, interpretability and accessibility.

In this project Data61 is focusing on existing indicators of tourism activity, such as overnight stays and spend, from a variety of sources. Therefore relevance and accessibility will not be considered. In a discussion paper on new indicators the ABS (2017) adds: scalable, sensitive, outcome focused and clarity. Of these, outcome focused and clarity are not relevant for the same reason. ORC International (2017) in their consideration of ADS and surveys, gave quality requirements of single source, accuracy, coverage and representativeness. ORC was looking at the set of indicators as a whole but the quality requirements are equally applicable to individual indicators.

Combining the ORC and ABS quality requirement relevant to ADS gives the following quality criteria, which are used to evaluate both traditional data sources and ADS in this report:

- Institutional environment (ABS): indicators should be selected from credible sources. Sources need to use consistent and accessible methodology.
- Accuracy (ABS): indicators should have a degree of precision to which the indicator correctly describes the phenomenon being measured.
- Latency or timeliness (ABS): indicators should have minimal delay between the reference period and the release of the data.
Coherence (ABS): indicators should be comparable across industries and presented as a time series. The latter is most important to tourism data consumers.

Interpretability (ABS): indicators should be accompanied with supporting information that can be understood by the general reader.

Scalable (ABS): indicators should be amenable to aggregation or disaggregation by topics of interest, e.g. industry, product, geography and time scales.

Linkable or single source (ORC): indicators should be either available from a single source or linkable so that relationships among variables can be examined.

Coverage (ORC): indicators should include all in-scope tourism activity. For example, all forms of expenditure (electronic, cash, prepayment) in a spend indicator.

Representative (ORC) (affects ABS Sensitivity): indicators should provide coverage of all the components of tourism activity so that no component is over- or under-represented. For example, all age segments of a population should be represented in the data. Sensitivity relates to the indicator being able to respond when the tourism activity changes. An unrepresentative indicator will not be sensitive to changes in those components not represented.

2.4.2 Institutional environment

All tourism data consumers saw ABS and TRA as trusted suppliers of tourism data. Both have the capability and track record to get the best out of the data and deliver quality products. While both have had technical issues with their data products or systems (troubles with census and NVS back-casting) the tourism data consumers are confident that both agencies are able to and do get on top of any issues. This project is indeed a request by tourism data consumers to have their trusted providers of data look at how new data sources might be able to service their current needs, knowing that due care will be taken with the results.

2.4.3 Accuracy

All tourism data consumers rated accuracy as the most important feature of tourism data, because the data are used to inform important decisions, such as multi-million dollar investments. Most viewed the IVS and NVS as extremely valuable products, being effective and consistent in recording progress against the Tourism 2020 targets.

IVS and NVS have the highest accuracy (as a percentage error) at the national level, declining by an order of magnitude at the state level and another order of magnitude at the regional level. The actual uncertainty varies at the finer spatial scales and roughly increases as the number of travellers visiting a spatial area decreases. Such a decline in accuracy, due to what statisticians refer to as “sampling error”, is unavoidable in surveys. Some tourism data consumers reported evidence of short term differences at regional levels between the IVS/NVS figures and what was observed or measured by other data sources. Some tourism data consumers interpreted this as indicating that the NVS/IVS figures were wrong, however the result may have fallen within the error bounds.

On the other hand, without an appreciation of technical issues such as sampling error and problems with the back-casting exercise required to adjust data to correct for bias, confidence in
the survey products’ applicability at finer scales may have been eroded. While this confidence might be restored through better education, the need remains for short term data at regional levels. This has contributed to the suggestion that ADS may be able to augment the survey products at these finer spatial scales.

The technique, employed by TRA, of rolling up multiple years of data to improve accuracy at the regional level may be unhelpful to marketing use cases, which often rely on information about short term trends. Another suggestion is that it may also be possible to use ADS to create more accurate results on very narrow data streams, such as visitor counts, which would in turn be able to help improve accuracy of the survey results at finer spatial scales. This assumes that visitor counts can be obtained with accuracy using ADS.

2.4.4 Latency or Timeliness

As mentioned above, the current quarterly releases of rolling annual IVS and NVS survey data, with a quarterly lag, is fit for national and state level reporting purposes. Tourism data consumers have generally built their processes around this data release schedule. In general tourism data consumers are happy with latency that is similar to temporal scale, i.e. monthly data provided within a month or weekly data provided within a week. Some felt getting data more often would significantly increase the cost of reporting, and would require significant changes to current reporting processes.

2.4.5 Coherence

Coherence depends on consistent definitions and methodology. The IVS and NVS are valued for their high coherence (with the noted one-time exception of the NVS back-casting exercise), which is ensured by the continuity of expertise in TRA and adherence to internationally accepted standards for collecting and processing tourism data.

On the other hand, the private sector sometimes presents their results as a black box. Without knowledge of the definitions and methodology used the coherence of the time series cannot be guaranteed and can be difficult to evaluate.

As evidence of the potential importance of coherence, some tourism data consumers reported that 5 – 10 years of data were required for their purposes. Longer data holdings would be valuable for analysing the impact on tourism of rare events (such as the GFC), since more such events would be captured.

Coherence is an important consideration in considering changing tourism data collection regimes and/or methodology. Ideally, when a new methodology (mainstream or ADS) is introduced a parallel run would be established to compare the difference in the time series with the old methodology.

2.4.6 Interpretability

The indicators collected in the surveys such as spend and accommodation nights are clear to all. However, what is less clear is the methodology that is used to ensure the values are accurate and representative. TRA is active in the Tourism Research Council (TRC) and assures Data61 that
researchers understand the methods used. However, their data consumers represent a wider audience than the TRC and might value a layperson’s explanation of statistical process.

### 2.4.7 Scalability

Most tourism data consumers would like tourism data with more accurate regional coverage. This was particularly true for those tourism regions outside capital cities.

As discussed in Section 2.4.3, the IVS and NVS have unavoidably high sampling error at smaller spatial scales. TRA addresses this issue by rolling up trends over multiple years. This practice introduces a delay into the release of data products. While this approach is suitable for reporting, and may be useful for some applications in supporting business, marketing is demanding more data at finer time scales.

Some ADS offer the potential for good scalability due to their high spatial and temporal detail, for example, geo-coded mobile phone usage data for a single carrier. Unfortunately such data are not always representative – but in some situations representativeness is not important, or can be improved if the limitations are understood.

The current temporal scale of the IVS and NVS data was seen by most to be acceptable, however of course some indicated they would be happy with more frequent releases if available, as a sort of 'wish list'. Some acknowledged the legitimate barriers and trade-offs involved in more frequent data releases, including cost, data quality, and comparison with other data releasing regimes. Again it was the marketing purposes and purposes of reporting to members that were cited as drivers for more timely data, such as monthly or even weekly releases. Some tourism data consumers said they would take data as often as it was made available.

To understand whether it is worthwhile to invest in producing data with finer temporal scale, it is important to clarify exactly what the real benefits would be.

### 2.4.8 Linkability

Data linkage is the process of associating individuals in one dataset with individuals in another, using a common set of attributes, or key. The idea is that two or more component datasets about the same or very similar groups can be linked to produce a single dataset with richer information about the individuals in that group. The IVS and NVS capture many attributes that are, by design, associated with the same individual.

A handful of tourism data consumers discussed the possibility of linking a range of tourism-related datasets to achieve richer data about the same individuals. However, there are many challenges to performing such data linkages including: additional privacy oversight, specialist expertise and software, as well as potentially the need to use a third party integrating authority. The Spanish study (section 2.3.3) showed it created a more valuable dataset.

All things being equal, linkability would be a desirable property for tourism-related data, particularly ADS, however currently it appears to be low in priority.
2.4.9 Coverage

As surveys ask individuals to recount their tourism activities, coverage is generally good. Some tourism data consumers questioned the ability of people to recall quantities (spend in particular) accurately. There were also questions about cultural sensitivities biasing some of the answers. TRA has assured Data61 that these have been taken into account in the design and execution of the surveys, via prompting and design learnings through pilot testing. There is a communication task required to make this clearer to tourism data consumers.

2.4.10 Representativeness

An accurate measure of national tourism activity and performance of the national tourism sector necessarily requires appropriate coverage of all components of tourism activity and the sector, with no component over- or under-represented.

The majority of tourism data consumers acknowledged that representativeness is one of the strengths of the IVS and NVS products, and importantly contributes to their accuracy. On the other hand, there was an appetite to further improve representativeness by finding a way to include, for example, access to airline lounges, and the cruise sector (despite the fact that it is quite small at less than 0.7% of the sector). Potentially, ADS could help with understanding how to account for such tourism components.

The back-casting exercise to correct representativeness bias inadvertently introduced into the NVS by the inclusion of mobile phones in the dialling pool has highlighted the whole issue of representativeness to tourism data consumers. It is recognised that ADS often have representativeness bias due to the underlying sample being restricted to a portion of the population (for example NAB covers only 14% of the credit card market). In some cases this can be compensated for, or at least explained.

To give a popular example, STR Global is a widely used data source, in the tourism sector, due to its good timeliness (low latency) and good metropolitan coverage. However, it only draws data from the hotels that have joined the service, most of which are 3 to 5 star. Thus, it has poor coverage of smaller accommodation providers that are disproportionately located in regional Australia.

For most purposes, poor representativeness can often be compensated for provided the limitations are well explained by data providers and well understood by users. Unfortunately there is evidence that this may not always occur, with caveats on data not receiving the attention they merit.

2.5 Sentiment of tourism data consumers

The sentiment about the IVS and NVS was generally positive. Virtually all tourism data consumers had high confidence in the IVS at national and state level. Confidence in the NVS was positive at the national level but at a state level some tourism data consumers were not seeing the accuracy they wanted. At regional level there were many complaints about accuracy. There were some demands for finer time scales so that the demand response to events or marketing could be clearly seen. There were also demands for more industry specific content such as conventions.
No one wanted to see the current survey products go. Many tourism data consumers were exploring ADS or were considering doing so. ADS appear to offer insight for at least short term, local activity, however no one claimed to know how they fitted in the bigger picture with the survey products. There were some calls to have trusted parties sort out the quality and applicability questions so that the tourism community could engage with the new data. Some agencies are already running trials but these trials are in the early stages.
3 Alternative data sources

An ADS is a data source which is not currently in normal use. That is there is no widespread acceptance as to how an ADS is used in tourism. The Department of Home Affairs and TRA are sharing both administrative and survey data in the building of IVS results and in replacing passenger cards. The administrative data is not an ADS because its current use is well understood. The distinction here between mainstream data sources and ADS is not the form of data capture (survey, administrative or personal) but the level of understanding of how to effectively use those data for tourism purposes.

For example, Data61’s consultations with state tourism offices found STR Global data is in such widespread use that it seems likely a consensus could be drawn on how to apply it for certain purposes. This consensus would not limit its use for other investigations but would allow all consumers to use it in a consistent way.

This section examines the use and utility of ADS. Section 3.1 looks at data use outside of tourism. ADS classes (section 3.2) are rated, in section 3.3, for the utility in providing statistics for types of tourism data. Finally, the risks of adopting ADS are considered in section 3.4.

3.1 Overview of data use in other industries

This section draws on two studies to examine data use outside of the tourism sector. The first is the Productivity Commission’s Inquiry into Data Availability and Use which examined impediments to data accessibility. The second is a study into potential use of ADS in employment.

3.1.1 Data availability and use

Yearly data generation is growing at an exponential rate according to the United Nations Economic Commission for Europe (as seen in the Productivity Commission report on Data Availability and Use 2017).
In 2017 the Productivity Commission (PC) released a report into data availability and use (Productivity Commission 2017). Their mandate was to look at the costs and benefits of making public and private sector data available and ready for use and propose changes in legislation and practice to improve them. They looked across both public and private data and data holders. Their recommendations are being reviewed by government with a response expected in early 2018.

What the PC found was that data frameworks and protections were not keeping pace with the rate of change of digital content. Their recommendations open up opportunities in many sectors including tourism. These opportunities are largely unknown until the data sources themselves are better known. This is true of ADS. The more they are available to trial and compare with existing approaches the clearer it will be as to how they might augment or replace the existing data sources. Some of the ADS may eventually carve their own niche for tourism reporting and analysis rather than compete with existing tourism data products.

The PC also found a lack of trust between data custodians and users in both government and the private sector. This lack of trust leads to numerous hurdles to sharing and releasing data, limiting its use and value for Australia. The need to rebuild this trust was a key part of the PC recommendations. While this trust was primarily focused on privacy and confidentiality there were also aspects of trust in the correct use of data and its correct interpretation.

Lack of trust, even in government data users, makes timely access to data difficult. Improvement in the flows of data inside government would improve the timeliness of TRA outputs. The PC proposed the creation of Accredited Release Authorities (ARAs) to streamline the release of data to users with different trust levels (public, government, project specific). These and the development of data and metadata publishing standards could dramatically improve the flows of data.

Coordinating the ARAs would be the National Data Custodian (NDC). The NDC would have the power to define National Interest Datasets (NID) which would support the collection and linking of data across all levels of government and could nationalise private sector data if necessary. It also
The Potential for Alternative Data Sources to meet Tourism Data Needs

proposed the opening up of Commonwealth Integrating Authorities (past ABS, AIHW and AIFS) to include more federal and also state agencies. The nationalisation of private sector data might drive down costs. The private sector could take the creation of NID as an opportunity to build trust with their customers in privacy preserving data handling and so open up that data to be linked with other datasets. Data61 and other organisations have made significant progress in privacy preserving data linkage. Two frameworks for de-identifying and sharing or releasing data have been recently published (O’Keefe, et al 2017 and Australian Computer Society 2017).

The PC also proposed a very powerful mechanism for data sharing, a comprehensive right to transfer data by individuals and small business. An individual would be able to request that an organisation or company holding data about them, transfer this data to a third party. This opens up an opportunity to recruit individuals to share data. The mechanisms for data exchange were seen as a private sector concern potentially facilitated by government.

3.1.2 ADS in employment

Data61 conducted a study of the use of ADS and modelling approaches in employment in 2017. Their report (Naughtin, et al 2017) examines trends for more real time sources of data and applying more complex (and computationally expensive) modelling techniques to the data. The mainstream data sources include surveys and administrative data combined by the ABS and other organisations to provide labour forces statistics and trends. The study found the strengths of mainstream data sources were:

- Representativeness of the labour market as a whole.
- Utility for long term forecasting.
- Capacity to be linked to other traditional data sources.

On the other hand:

- They can lag behind current trends in the labour market.
- They do not provide granular data on skills or specific populations.
- Privacy or accessibility can limit data linkages.

When looking at ADS, the most common form in this domain being online job advertisements, it found that their strengths were:

- Timely and up to date insights.
- Identifying current and emerging labour market trends.
- Capacity for fine grained analysis.

With weaknesses of:

- Lack of capacity for long term projections due to short term variability.
- Biases which limit its representativeness (online advertisements favour professional jobs).
- Might not be accurate (advertisements may simply be sounding out the market).
- Unstructured online data lacks consistency leading to missing or duplicate data.
Despite these limitations the strengths were enough to spark new data analyses, models and companies in Australia and overseas. They did not replace the mainstream employment products but augmented them because they served a need that was not met. For example, Burning Glass Technologies has developed a Labour Insight Tool that draws online job advertisements from over 40,000 job boards (worldwide). It does this automatically, providing an update every two days. It is used across government, education and recruitment agencies for career services, course design and marketing purposes.

Another example is the Australian Government’s Labour Market Information Portal. It uses mainstream employment products to provide employment projections. Alongside this it publishes the Internet Vacancy Index (IVI²), drawn from online job portals to provide monthly insights. By labelling it IVI it declares it as an internet specific indicator without the backing of the ABS.

The report goes on to seek to bring best practice together in an Australian portal and proposes to explore ways of combining and linking many of the approaches in an Australian context. The diagram below shows the current employment data and modelling landscape.

![Figure 3 Current employment data ecosystem.](image)

### 3.2 Alternative data source classes

The tourism data consumer consultations identified a large number of ADS, the list of ADS was also grouped into classes of related datasets. The benefit of packaging the ADS in terms of these

---

classes is that it reduces the number of data sources to be considered to a manageable size while maintaining the overall relevant characteristics of the collection of ADS in the class. For example, when considering social media as an ADS, different social media platforms can be considered collectively for their suitability as an alternate source of tourism data.

The classes decided upon are:

- Transactions
- Social media
- Booking/distribution/ticketing
- Telecommunications

One class of ADS that has not been explored in this report is the data collected by tourism industry bodies about their members (such as hotels, restaurants, conventions, etc.). These are samples of participants in those industries and the data collected is often local, rich and timely. More could be done to coordinate the data resources from across these bodies to provide a substantial and national ADS.

The following sections provide an overview of these classes, describing how they may be used for tourism purposes.

### 3.2.1 Transactions

Transaction data is often held up as a powerful source of tourism information. Australian Payments Clearing Association Limited (APCA) data suggests Australians made 8 billion transactions through credit, debit and charge cards last year and the number is growing (APCA 2017). Every time a credit card, debit card or EFTPOS is used a wealth of data is recorded:

- Card number – sometimes linked with card owner.
- Merchant identity – their name may reveal the type of business as well as location.
- Merchant location – as provided by the merchant.
- Transaction value.
- Transactions class - by a four digit international code.
- Date and time.

In theory this can describe who, what, where, when and how much is spent. In practice it is more complex and this will be explored further in section 3.3. Card providers must protect the privacy of their customers, however new technologies are making this easier. Data aggregators and secure infrastructure providers are able to link datasets in the private sector and provide secure access to de-identified and aggregated data.

Individual transaction providers such as the Commonwealth Bank of Australia (CBA) provide business insights to customers based on transactions through their merchant point of sale systems. For example, they provide Daily IQ which tells businesses about their turnover and those of similar businesses in their area – highlighting that such data sources can be used at small spatial and temporal scales. Preliminary investigations into Citibank and HSBC found extensive use of their own transaction data, however there were no clear data products for sale.

---

3 Interpretations, or conclusions formed on the basis, of the data or information, are entirely CSIRO’s.
3.2.2 Social Media

Social media monitoring involves the targeted monitoring of various social media channels focusing on specific topics, products, brands, competitors or cohorts depending on user or business needs. It is possible to capture user locations, activities or interests and analyse the data to gain an insight into current attitudes, sentiment or behaviours and trends over time. These insights can be actionable, facilitating intervention as needed.

There are numerous free and commercial social media monitoring tools available that focus on specific social media platforms such as Twitter, Reddit, Blog sites, Facebook, Instagram, YouTube or Instagram. Social media monitoring can also focus on specific application areas, for example emergency management, law enforcement, counter-terrorism, financial and insurance services, media, marketing, brand management, popular culture, customer engagement and so on. More sophisticated commercial offerings provide an extensive monitoring infrastructure across numerous social media platforms. There are research prototypes for effectively identifying content of interest, analysis and summary of content, understanding the network of communication and effective big data visualisation.

The social media vendors are also leveraging their social communities by offering commercial arrangements for advanced access to their platforms to help managers of social media accounts understand how the online community are engaging. Notable examples are Facebook Insights, Twitter Analytics, YouTube Analytics and Google Analytics.

3.2.3 Booking/Distribution/Ticketing

For many tourist activities, a booking or ticketing process is essential; this includes items such as accommodation reservations, event or tour bookings, transportation ticketing, airline or car rental bookings etc. All of these activities provide data points from which valuable tourism information may be extracted.

In the early stages of online reservation systems, Global Distribution Systems (GDSs) were designed to connect travel vendors to travel agencies and consequently the consumers. Although the hotel and car rental industries continue to benefit from these systems for last minute bookings, many vendors (in particular airlines) developed their own online ticketing systems.

Furthermore, online travel agencies (OTAs), which include websites such as Webjet.com.au and Booking.com, are now major platforms for travel bookings, as consumers have tended to move away from in-person travel agencies (Carey, et al 2012). OTAs provide accommodation, airline, car rental, and tour package bookings and, therefore, have the potential to provide tourism data across a range of sectors. Additionally, Airbnb has enabled individual homeowners to be able to advertise rooms and properties for short and medium term rent directly through their online platform. This has further expanded the accommodation market from traditional hotels and bed and breakfasts.

Another source of electronic ticketing is metropolitan transportation systems which have evolved to use swipe on/swipe off cards rather than paper tickets and these systems also have the potential to capture tourism activities. Although the majority of use on these systems is not tourist
related activities (residents commuting for example) there is potential for tourism data to be extracted from this information.

Merchant booking management systems (what merchants use to manage their product use) are the flip side of ticketing. ADS such as STR Global and V3 TXA collect data from the merchants themselves. The latter uses widespread V3 accommodation management software to facilitate this.

3.2.4 Telecommunications

Personal electronic devices are ubiquitous and considered an essential item in the modern world. People expect to be able to stay connected at all times and this is equally true when travelling. Mobile phones and other devices are used to stay in touch with family and friends, to record experiences and to keep informed of world events. While the means to facilitate this are generally smart phone apps or access to a web browser, the underlying technology needed is the telecommunications infrastructure.

The telecommunications industry can provide tourism insight by monitoring the activity of their customers as they interact with the telecommunications infrastructure such as mobile phone towers or through Wi-Fi connections. An individual will have typical behaviour centred on home and work life which can be distinguished from atypical activity which can be considered tourist activity. This activity can be phone calls, text messages, internet usage, ‘pings’ or GPS locations. The information collected is not the content of the interaction, but the interaction itself: when it occurred, for how long and the location of the equipment used to facilitate the activity.

Telecommunications companies have detailed information about their post-paid customers and this can be de-identified and aggregated to conform to privacy legislation and community expectations. Detailed information exists about the movement of prepaid customers however their context (such as home address) is not necessarily held which would make it harder to determine their tourism behaviour.

In addition to phones, location is captured through devices connecting to Wi-Fi hotspots, GPS tracking of fitness devices, GPS location through apps such as Google maps and transportable Internet of Things (IoT) devices as they move around the IoT\(^4\) network. The IoT network is being trialled in Australia, one form of this network is LoRaNet.

3.3 Alternative data sources to inform tourism statistics

An ADS provides an alternative source of insight for tourism data consumers. In order to evaluate the utility of ADS for tourism statistics Data61 first classified the data for tourism statistics into five types as described in section 2.1.2: Accommodation, spend, movement, purpose of visit and sector impact. They provide a high level categorisation of the areas of interest for tourism data consumers and reflect the information currently obtained from the survey products. The purpose of this categorisation is to provide a framework to analyse classes of ADS for their utility to tourism.

\(^4\) http://www.iot.org.au/
The classification of tourism data requirements into types simplifies this task.

Second, Data61 examined ADS from each of the four classes described in section 3.2. The full list of ADS Data61 drew from has been summarised in additional reference material provided to Austrade and Tourism Australia. This list was obtained from interviews with tourism data consumers and refined in discussion with TRA and TA. From this list, a major ADS from each class was chosen by Data61. These providers were interviewed, or if unavailable, their website researched for ADS content. The major ADS providers examined were: Quantium\(^5\) (transactions), Twitter (social media), AirDNA (Booking/Distribution/Ticketing) and Optus (telecommunication). Data61 does not attempt to make any inferences as to whether one ADS from a class is better or worse than another in the same class. In addition, the web sites of some other ADS in each class were examined for their ADS content, as time permitted, in order to provide a broader view of ADS availability.

Most ADS providers have not focused on tourism as the primary use of their data, given the data’s multi-sectoral possible uses and often core business focus i.e. bank data is used to inform bank decisions first and foremost.

Sections 3.3.1 to 3.3.4 contain the analysis of ADS classes, for each of the tourism data types, against the tourism data quality criteria described in section 2.4.1, namely: institutional environment, accuracy, latency or timeliness, coherence, interpretability, scalability, linkability, coverage and representativeness. These are the same criteria used to assess the mainstream data products in section 2.4. Table 2 below provides a qualitative rating for a data set’s utility derived from the ADS analysis. It is designed to show the strengths and weaknesses of individual ADS classes in providing data useful for tourism statistics in each of the tourism data types.

<table>
<thead>
<tr>
<th></th>
<th>Accommodation</th>
<th>Spend</th>
<th>Movement</th>
<th>Purpose</th>
<th>Impact</th>
</tr>
</thead>
<tbody>
<tr>
<td>Transactions</td>
<td>Low</td>
<td>High</td>
<td>Medium</td>
<td>Low</td>
<td></td>
</tr>
<tr>
<td>Social Media</td>
<td>Low</td>
<td>Medium</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Booking/Distribution/Ticketing</td>
<td>High</td>
<td>Low</td>
<td>Medium</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Telecommunications</td>
<td>Low</td>
<td>Medium</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2: Utility of Tourism Data Source Classes in each of the types of tourism data

Each ADS class is ranked as High/Medium/Low which corresponds to Data61’s judgement of the utility of the ADS class.

- ‘High’ means that the data source is well suited to provide insight into the tourism data type and is considered a strong candidate for further examination.
- ‘Medium’ means that the data source can provide some general or indicative insight.
- ‘Low’ means the data source is of limited use.
- An empty cell specifies that the data source cannot be used to infer information about the associated tourism data type.

\(^5\) Many tourism data consumers had been approached by Westpac. A different transaction provider was chosen to provide additional insight.
The existing mainstream tourism data products are high quality in all tourism data types. In this qualitative assessment Data61 is not directly comparing ADS against the mainstream products. That is best done in the context of the purpose and scale of the data used and when both sets of data are available for comparison. These findings should be further explored through deployment and review of ADS services as described in the Roadmap in Section 4.

3.3.1 Transactions

Transactions are a high volume and high speed data source. Accounting systems such as Xero and MYOB receive daily data feeds from banks to update their systems with transactions from their customers. Only digital transactions (credit cards, debit cards, EFTPOS, stored value cards and ATM withdrawals) are covered in transaction data. The whole cash economy is missing. The Reserve Bank of Australia's 2016 Consumer Payments Survey (Doyle et al 2017) estimates cash at 40% of all payments (by number of payments). Applied to tourism, this could be a significant amount of the overall spend. Work in NZ found that use of cash by tourists industry varied widely (from 20% to 80%) with country of origin making it harder to compensate for (NZ 2012).

Data61 spoke with Quantium about transaction data. They hold NAB transaction data as well as information about the NAB customer base. Data61 heard that some other transaction data providers could provide data for any card used through a point of sale device. This may potentially capture a larger number of transactions but at the loss of being able to link the identity of the card owner to other information about them such as address, age and family information. Without this additional information the ‘home’ locations of a card holder need to be derived from the transaction record, for example by looking for regular visits to supermarkets. International cards may have overseas designations, this has not been investigated. Therefore in some cases a transaction data source will most likely be relevant to national travellers, and might treat long stay international travellers (who create local bank accounts) as national travellers.

China shows growing use of mobile devices for payments. The Chinese Internet Network Information Centre (CNNIC 2017) estimates in a 2016 survey that 53.2% of Chinese are using the internet and 95% of the most commonly used mobile devices to access it. Some 67.5% of internet users also make some mobile payments (up by ~30%) and 50.3% of internet users pay bills through mobile payments when shopping at a physical store.

Quantium links their credit card data with Foxtel, Woolworths and CoreLogic (real estate) data, though these data source are less useful for tourism.

Accommodation use is determined through spend on accommodation services. There is a merchant code (a four digit identifier) which is tied to the class of service for which a merchant is registered. Some codes are specific to large accommodation vendors, while other vendors may be assigned to a general code for ‘Lodging – Hotels, Motels, Resorts, Central Reservation Services (not elsewhere classified)’. Note that a hotel with a restaurant may use the same code and therefore confuse the type of spend (VISA 2017).
With investment in the dataset, a merchant code, name and location can be used to identify specific business and the class of accommodation. This sort of investment is not trivial and is ongoing, representing high value add in the data. It is the consistency and quality of this value add that exemplifies a strong institutional environment. Data61’s assessment is that this institutional environment is likely to be variable across transaction data providers.

Prepayments

Prepayment of travel can account for a large proportion of travel spend. If these transactions are missing from the data then the type of spend will also be missing. TRA has provided Table 3 which covers the international prepaid trip spend in 2016-17 by major country of origin, from the IVS.

<table>
<thead>
<tr>
<th>PURPOSE OF VISIT/COUNTRY</th>
<th>NEW ZEALAND (%)</th>
<th>JAPAN (%)</th>
<th>UNITED STATES OF AMERICA (%)</th>
<th>CHINA (%)</th>
<th>UNITED KINGDOM (%)</th>
<th>OTHER (%)</th>
<th>TOTAL (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Holiday</td>
<td>47</td>
<td>63</td>
<td>68</td>
<td>61</td>
<td>51</td>
<td>48</td>
<td>54</td>
</tr>
<tr>
<td>Visiting friends and relatives</td>
<td>44</td>
<td>58</td>
<td>60</td>
<td>26</td>
<td>56</td>
<td>50</td>
<td>47</td>
</tr>
<tr>
<td>Business</td>
<td>53</td>
<td>64</td>
<td>68</td>
<td>41</td>
<td>64</td>
<td>57</td>
<td>58</td>
</tr>
<tr>
<td>Employment</td>
<td>13</td>
<td>24</td>
<td>21</td>
<td>6</td>
<td>11</td>
<td>11</td>
<td>12</td>
</tr>
<tr>
<td>Education</td>
<td>37</td>
<td>37</td>
<td>56</td>
<td>18</td>
<td>25</td>
<td>27</td>
<td>24</td>
</tr>
<tr>
<td>Other reason</td>
<td>56</td>
<td>54</td>
<td>82</td>
<td>36</td>
<td>55</td>
<td>35</td>
<td>43</td>
</tr>
<tr>
<td>Total</td>
<td>44</td>
<td>57</td>
<td>64</td>
<td>32</td>
<td>49</td>
<td>40</td>
<td>42</td>
</tr>
</tbody>
</table>

Table 3 Percentage (%) of trip cost prepaid by country of origin and purpose 16/17 (TRA)

Similarly for domestic overnight travel significant proportions were pre or post-paid. Table 4 below shows proportion paid during a trip, before/after a trip, or by a third party.

<table>
<thead>
<tr>
<th>PURPOSE/PAYMENT TYPE</th>
<th>PAID DURING TRIP (%)</th>
<th>PAID BEFORE/AFTER TRIP (%)</th>
<th>PAID BY EMPLOYER/OTHER NOT TRAVELLING (%)</th>
<th>TOTAL (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Holiday</td>
<td>68</td>
<td>31</td>
<td>1</td>
<td>100</td>
</tr>
<tr>
<td>Visiting friends and relatives</td>
<td>73</td>
<td>26</td>
<td>1</td>
<td>100</td>
</tr>
<tr>
<td>Business</td>
<td>44</td>
<td>20</td>
<td>35</td>
<td>100</td>
</tr>
<tr>
<td>Other reason</td>
<td>69</td>
<td>25</td>
<td>6</td>
<td>100</td>
</tr>
<tr>
<td>Total</td>
<td>63</td>
<td>27</td>
<td>10</td>
<td>100</td>
</tr>
</tbody>
</table>

Table 4  Payment timing for overnight domestic trips 16/17 (TRA)

Accommodation data is primarily about number of nights spent in a location. The accuracy of this depends on obtaining a figure for spend on accommodation at a location and transactions at
different locations on days either side of the stay to confirm the duration. It is highly dependent on the temporal spread and frequency of transactions. With cash payments unseen this may lead to significant errors in the number of nights. Prepaid accommodation is also problematic, especially if payments go through a travel agent rather than directly to the merchant. Prepayments with the merchant can be mapped onto the right time when travel is detected in the location. Errors in locations of merchants can also be corrected for by cross-referencing possible accommodation locations and travel.

Explaining how the number of nights of accommodation was determined could be difficult so the interpretability of the result may not be robust. Spatial and temporal scale depend on the size of the sample. NAB data represents about 14% of the credit card market. However privacy concerns may require low counts (<10 up to <50) to be aggregated. Uncertainty in the location of accommodation could also see that count aggregated up spatially. The difficulty in judging the number of nights makes transactions data of low use for Accommodation.

Spend is perhaps the biggest strength of transaction data. The merchant code does provide an easily accessible breakdown on spend. Cross-referencing the merchant ID with a more detailed description of the merchant can make it much richer. However, as previously mentioned, the coverage is not good. Not only are cash transactions omitted but so are electronic transactions from other card providers and many prepayments. It is also unclear whether one bank's customers spend in the same way as another bank's while travelling. Comparison of tourism data derived from different transaction sources would be insightful, to enable confidence in the data and removal of bias in the sample.

Estimates can be made on the cash spend while travelling. Monitoring ATM and EFTPOS withdrawals over a longer journey can indicate cash spent. ABS retail spend statistics can also be used to calibrate these, but without a full knowledge of cash spend there is a significant gap in the total spend amount. However, just knowing electronic spend may be useful in itself. Despite all this, transaction data are still high in their applicability to the Spend category.

Measuring movement is based on merchant location in transactions. Errors in merchant location can be compensated for by looking at location from temporally nearby transactions. However, the main issue with location is if nothing is spent electronically then no location is recorded. Still major stops are likely to have transactions so its utility to movement is medium.

Purpose is somewhat implied by the type of things money is spent on but it is by no means clear. Its utility is low. Sectorial impact is difficult to gauge from transaction data. Quantum’s other data such as real estate may give a better indication, however the linked nature (a single person who travels) of the real estate data is not valuable for tourism. Impact does not rate well enough to list in the table.

With accounting firms receiving daily transaction feeds, latency and temporal scale can be very good. However, that depends on the quality of the data. New merchants, not seen before by the ADS provider, will have little context or metadata. Estimates where they are involved will be lower quality until that metadata can be built up. Determining location and accommodation use can require multiple days of data, but with automated processes latency can be low and temporal scale in the order of weeks to a month.
Credit card and cash transactions

In the Reserve Bank’s 2016 Consumer Payments Survey, cash use continues to decline as a proportion of the total number of payments made under $10 000.

Figure 4 Proportions of number of transactions by transaction type (Doyle, et al 2017)

For credit card transactions the major banks dominate the market (Table 5).

<table>
<thead>
<tr>
<th>CREDIT CARD PROVIDER</th>
<th>MARKET SHARE (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Commonwealth Bank</td>
<td>27</td>
</tr>
<tr>
<td>Westpac</td>
<td>23</td>
</tr>
<tr>
<td>ANZ</td>
<td>19</td>
</tr>
<tr>
<td>National Australia Bank</td>
<td>14</td>
</tr>
<tr>
<td>Citigroup</td>
<td>12</td>
</tr>
</tbody>
</table>

Table 5 Market share of credit card providers by size of card balances (APRA 2017)

Coherence of the data is a challenge. The NZ studies showed that the use of credit cards (compared with cash) varies over time and with different spatial regions and country of origin. These factors proved difficult to compensate for in NZ. Without this compensation a time series of transaction data is not very meaningful. This also speaks to the representativeness of transaction data. Is the NAB data representative of all transactions? Quantum uses ABS population segments from the census to scale up the transaction and calibrate it against retail trade data in large population centres. However, there may also be some spatial changes in the use of electronic transactions and the use of cash, especially in digitally remote areas.
The biggest downside of transaction data is the lack of a single source. Transaction data could be much more effective if it was linked with telecommunication services (as in the Spanish study, section 2.3.3), booking data from online travel agents, possibly hotel ticketing information and a conversation with travellers about purpose, number of travellers and use of cash.

### 3.3.2 Social Media

Social media is in wide use by the general public and is now seen as a standard communication channel. The community expects to be able to interact online with government and business and this includes social media interactions. Some government agencies are actively monitoring content on social media to engage with customers and clients, gauge community sentiment, obtain feedback about policy changes and government discourse. For example, recent contentious issues have been the Centrelink ‘not my debt’ campaign and ‘robo-debt’ initiative, the NSW Government greyhound racing ban, the ‘lock the gate’ action group and community backlash against mining activities such as fracking and coal seam gas.

The ability of social media monitoring tools to identify target groups in order to engage with them is well known. This can be through targeted advertising on Facebook or Twitter, customised offerings through rewards programs to influence shopping behaviour, or more recently to reinforce belief systems via the echo chamber effect where people are presented with online content that conforms to their world view.

This ability to monitor social media can be used to gain an insight into user behaviour and opinions. The challenging task is to identify, or at least approximate, the relevant cohort present on the social media platform, to reasonably estimate their location to appropriate geographic scales, characterise their behaviour and attitudes, aggregate the results and track these characteristics over time. Solving these issues are active areas of investigation by the research community and are under continuous development by tool providers. Note, that if social media monitoring tools are used then care would need to be taken to mitigate public perceptions of mass surveillance.

Instead of monitoring social media using third party tools, another option is to collaborate with the social media provider directly. For example, Facebook profile their users for purposes of targeted advertising. It is expected, although not confirmed, that Facebook could identify which of its Australian users are planning a future holiday, who is currently on holiday and who has recently been. By identifying these different user cohorts, their Facebook content can be processed to determine where they went, to gauge their experience as being positive or not and the reaction of their friendship network. Note, that this information should be able to be aggregated by Facebook to preserve individual privacy. Similar engagements would need to be undertaken with the other major social media platforms.

In general, social media is seen as a credible source of information although there are anomalies particular to the environment to be mindful of, such as ‘fake news’, ‘bot’ activity, trolls, spam and general misinformation. In terms of the quality framework for assessing social media to inform tourism data, the main benefit of social media is its timeliness; it is a good source of real-time content that can be tracked as time series data. Other benefits are its interpretability as content can be directly obtained which leads to gaining an insight into the sentiment of the user. It can be
scalable depending on the geographic resolution of the available content, which varies across and within platforms. However, the information available has low accuracy and coherence for the reasons outlined above.

### Social media use

The annual Sensis survey of social media use in Australia (Sensis 2017) reports how consumers and businesses use social networking sites. The June 2017 results show that about 80% of people use some form of social media platform with about a third using it more than five times a day. Facebook is the most popular platform; 94% of Australian social media users are on Facebook, 46% on Instagram, 40% on Snapchat, 32% on Twitter and 18% on LinkedIn.

Almost all 18-29 years olds are on social media (99%) and this use decreases with age, but initially only slightly. For the 30-39 age bracket 96% use social media; 40-49 it is 86%; 50-64 66% and for the 65 year olds and over it is 47%. It is expected that the percentages in the older age groups will increase over time as people age and continue to use social media.

Roy Morgan (2015) reports Twitter usage (Figure 5) skewed towards the younger population and capital cities.

![Figure 5 Twitter usage in Australia based on a sample of people aged 14+ (Roy Morgan 2015)](image)

Cowling (2017) reports Facebooks count of active users by age category (Table 6). In the same report Cowling notes “A significant proportion of the Chinese community in Australia use Chinese social media tools like WeChat, RenRen and Weibo and these networks are growing”.

38 | The Potential for Alternative Data Sources to meet Tourism Data Needs
According to the Chinese Internet Network Information Centre (CNNIC 2017) a 2016 survey lists the three major Chinese social media sites as WeChat Moments 85.8% of internet users, Qzone 67.8% and Weibo 37.1%. WeChat is a relatively private site allowing sharing of content between friends. Weibo is an open comment site (similar to Twitter) and Qzone is somewhere in between.

<table>
<thead>
<tr>
<th>AGE BRACKET</th>
<th>NUMBER OF USERS</th>
</tr>
</thead>
<tbody>
<tr>
<td>13-17</td>
<td>940 000</td>
</tr>
<tr>
<td>18-25</td>
<td>3 500 000</td>
</tr>
<tr>
<td>25-39</td>
<td>6 100 000</td>
</tr>
<tr>
<td>40-55</td>
<td>4 100 000</td>
</tr>
<tr>
<td>55-64</td>
<td>1 600 000</td>
</tr>
<tr>
<td>65+</td>
<td>1 200 000</td>
</tr>
</tbody>
</table>

Table 6 Count of active Australian Facebook users (Cowling 2017)

Each social media platform has bias towards different demographics. The social media platform would need to adjust the data based on demographics to ensure it is representative. Also, if each platform is providing aggregated results then the coverage is unclear. People do use multiple social media platforms and their information may be counted twice if a platform’s aggregate data is further combined with data from other platforms. For these reasons it has been rated as ‘Low’ for movement and ‘Medium’ for purpose as can be seen in Table 2.

3.3.3 Booking/Distribution/Ticketing

Due to the broad nature of this ADS class, booking, distribution, and ticketing systems cover a range of the tourism data types. Accommodation data could be primarily sourced from STR Global, which is widely used amongst tourism data consumers, or V3 TXA or OTAs. The Airbnb market could be covered by data supplied by AirDNA.

Sources such as STR Global have demonstrated that they are reliable and credible sources and therefore they have a strong institutional environment. In contrast, AirDNA is newer and due to its current data collection method of scraping the Airbnb website and the fact that it has no formal agreement with Airbnb, it is potentially a less institutionally stable data source.

The accommodation data products tend to have low latency and can be reported on monthly or weekly timescales. Interpretability is also predicted to be high as raw hotel availability and bookings are typically easily understood by tourism data consumers. A major concern for STR Global is that they only cover hotels of 10 rooms or greater and tend to only represent the larger hotel operators. This further impacts the geographical coverage as larger hotels tend to be in more metropolitan locations and regional areas do not have high coverage from this data source.

Tourism data consumers have also reported that coverage in Western Australia and Tasmania is particularly low.
AirDNA is able to provide data on all Airbnb sites listed in Australia and therefore is entirely representative of the Airbnb market. The location of these listed rooms tend to be in metropolitan or popular tourism areas, however many tourists also turn to the Airbnb market in order to be able to stay in areas that have a low supply of hotel accommodation. Thus, in addition to traditional hotel data, Airbnb data may provide additional insights into where tourists are staying.

Airbnb lists individual saleable units (rooms, houses, apartments) even if they exist in a block of units such as a bed and breakfast or lodge. As the unit transitions to ‘not available’ AirDNA records it as ‘booked’ with the last price. Rooms can be marked as either ‘blocked’, when the person listing the room specifies that it is not available, or ‘booked’ when the room is booked. Airbnb withdrew this information recently marking listings simply as unavailable. AirDNA is not able to distinguish between these two settings but relies on the proportions from historical data that it had obtained when the information was available. This is likely to raise some accuracy issues, however aside from this issue it covers bookings on Airbnb.

**Travel agents market share**

With such a large spread (Figure 6) of OTA market share the challenge is to bring a reasonable number of them "to the table" to provide data.

![Figure 6 Percentage of Australians who used a travel agent in 2016 (Roy Morgan 2017)](image)

Amongst GDS there are three main players controlling flight bookings: Amadeus, Sabre and Travelport. Business Travel iQ (2017) reports their market share in Figure 7.
Coverage in accommodation is helped by STR Global’s exclusive licencing. It means their data can be added to data from another source. With the other ADS it is less clear. AirDNA infers bookings from changes to Airbnb’s web site. As properties are multi-listed across other OTAs, data from other OTAs cannot be simply combined. The Australian Hotel Association also collects booking data from its members on a monthly basis. This might be a potential "other source" of accommodation data. Many OTAs request information (number, age, billing address) about the guest’s booking but this is not accessible to AirDNA, due to its weak relationship with Airbnb. This would make it hard to link AirDNA’s data with data from OTAs.

With enough coverage, accommodation data is likely to be representative. Individual sources such as Airbnb may favour those more familiar with internet booking systems and this proportion is likely to change over time, affecting coherence.

In addition to the accommodation market, a medium level of utility is predicted from booking, distribution, and ticketing sources for the movement data type, in particular for airline travel. Global Distribution Systems (GDS) as well as established OTAs such as Booking.com would be considered highly reliable sources with high institutional stability. Timeliness of this type of data is also good as it appears to be available on a monthly basis. This data would only cover bookings made through such websites or systems and therefore would not cover the entire market, in particular bookings made directly through the airline’s website. Some movement can be determined from the chain of accommodation bookings, but it is dependent on good accommodation coverage. TRA is exploring access to GDS data.

Given that air travel capacity data is currently available from BITRE, these data sources could be used to validate and calibrate any airline booking data from alternative data sources. This ability to correlate with existing sources would be highly important. With this in mind, the overall utility of this source for movement would still be relatively high.
All of the sources mentioned above also contribute data to the Spend data category. However, they would only provide data on spend in the accommodation or movement sector which is likely to be only a portion of the total spend by a tourist on any trip. Food, gifts, or other purchases for example would not be included in the estimate of spend. The GDSs, OTAs, as well as AirDNA all provide the amount the room or travel was booked for as part of the data provided and the same representativeness, accuracy and institutional environment issues raised previously would apply. Due to the low coverage, Data61 have rated this to have Low utility for the Spend category.

### 3.3.4 Telecommunications

The main benefit of telecommunications data is the insight it can provide into the volume and location of devices and by inference people. This information can be processed to distinguish between locals and visitors but it does not provide an insight into the purpose of the visit, if they enjoyed it nor how much they spend while there. These aspects can be estimated by combining other ADS and there are research efforts underway that aim to explore the feasibility of this approach, referred to as Tourism Big Data.

For example, (Heerschap et al 2014) present three case studies of using big data for tourism statistics: using web crawlers to compile tourist accommodation lists; using log data from a smart phone app to monitor tourist activity; and anonymous aggregate mobile phone metadata based on Call Detail Records from collaborating phone companies (Vodafone and Mezuro) to also monitor tourist activity.

Similarly, (Ahas et al 2014) investigated if passive mobile positioning data from mobile network operators (MNOs) could be used by national statistical institutes to produce official tourism statistics by concentrating on the legal, business and technological aspects of the task within the European Union. They found that there is value in the data available to MNOs and they are seeking to utilise this data commercially but there are concerns in terms of regulation and public opinion.

Also, (Song and Liu 2017) investigated repositories of big data relevant to the tourism industry to inform evidence based decision making. This data includes web and social media, sensing devices, telecom network data and financial transactions. They note that tourism big data has advantages over traditional survey methods: reliability, new information flows and real-time data and nowcasting. They note a number of possibilities and benefits of big data for tourism outcomes and describe a proposed framework for tourism forecasting. The implementation of their recommendations is left as future work.

Other references of interest include (Wong et al 2017, Höpken et al, Shafiee and Ghatar 2016) as well as various online references6.

While telecommunications can tell where someone visited they cannot tell how many other people without devices were present. For example a young family may have no devices associated

---

http://www.collectivesensing.org/tourism-in-italy/
with children. Conversely, the parents may have multiple devices which may give multiple counts. As with most ADS, the quality and breadth of information is improved as data is linked.

As with credit cards, post-paid phone accounts provide details of the account holder and therefore more context. International tourists can be identified by international roaming phones, however as roaming calls are expensive they are less likely to use it and therefore give less location data. For this reason tourists do purchase prepaid sim cards. Distinguishing tourism through prepaid sim cards is challenging. For national tourists it requires identifying long term local behaviour (such as regular calls made from a shopping centre) to identify a ‘home’ for the card holder. International tourists might be distinguished by the brevity of their activity in Australia. The accuracy of these measures are yet to be determined but clearly add complexity.

A single telecommunication provider will only represent a portion of the population, mechanisms would be needed in order to scale up the coverage. Different providers are known to represent different portions of the population, which brings representativeness into question. Finally, some locations are serviced by only a single mobile telecommunication provider. In those instances a fine spatial scale is not feasible and must be aggregated spatially until reliable data is available.

By merging data from across telecommunication providers, better representativeness and coverage can be attained. However there are potential commercial constraints to this approach. Furthermore, as company market share changes over time, the coherence of the data may change. Data will subsequently need to be re-calibrated to compensate for these changes.

Australian telecommunications

The Australian telecommunications landscape is dominated by three major network operators as shown in Table 7. Not only do they have significantly different market share they also have different coverage. The differences in population coverage may be small but the last few percent of the population are in regional and remote areas and lead to large differences in landmass coverage.

<table>
<thead>
<tr>
<th>PROVIDER</th>
<th>MARKET SHARE (%)</th>
<th>POPULATION COVERAGE (%)</th>
<th>LANDMASS COVERAGE (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Telstra</td>
<td>45</td>
<td>99.3</td>
<td>&gt;31</td>
</tr>
<tr>
<td>Optus</td>
<td>~25</td>
<td>98.5</td>
<td>15.6</td>
</tr>
<tr>
<td>Vodafone</td>
<td>~20</td>
<td>96.0</td>
<td>7.5</td>
</tr>
<tr>
<td>Other</td>
<td>10</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 7 Mobile telecommunications market share and coverage (Productivity Commission 2017a)

According to ACMA (2016) the number of mobile services grew 2.6% across 2015-2016 but this increase has been variable over the years. They also note that telecommunication operators are committed to 5G networks by 2020 and exploring voice calls over 4G and 5G networks.

The Productivity Commission (2017) looked into the data collection capabilities of wearables. Over 2 million Australians possess a wearable device. Fitbit disclose the potential sale of de-identified
tracking data to third parties on their web site. This is a potential source of data from both national and international visitors. Fitbit know a lot about individuals but its representativeness is unclear.

For these reasons, the use of telecommunications content to inform tourism data has been rated as ‘Low’ for accommodation and ‘Medium’ for movement as can be seen in Table 2 above. Telecommunications content is a credible source of information about the location of people. It can be used with high accuracy to locate people identified as being likely tourists, although it may be difficult to exclude seasonal or fly-in-fly-out workers or international visitors who use prepaid SIMs; although they would be identified as a visitor or tourist, they could not easily be identified as an international one. This information can be made available in near real-time that can be tracked as time series data and is scalable depending on the geographic resolution of the aggregate information made available.

3.4 Risks of using ADS

The New Zealand experience (described briefly in section 2.3.1) of using ADS (credit card data specifically) to develop Regional Tourism Indicators (RTIs), that were later replaced by Provisional Regional Tourism Estimates (PRTEs), is insightful of some of the risks. NZ captured transactions from local Bank New Zealand (BNZ) customers and transactions of foreign credit cards through local machines. They worked with a private data specialist, MarketView, with government assistance and supervision.

Initially NZ produced RTIs, a number tied back to 2008, on a monthly basis. This was to help users look at movements (particularly larger movements) rather than trying to interpret dollars. A Regional Tourism Estimate (dollar value) was produced annually where there was capacity to weight it with other data. The users had difficulty interpreting the RTI indices and in 2015 it was recommended to replace the RTIs by PRTEs that included a dollar amount. The recommendation highlighted the risks to be managed (NZ 2015). Namely:

- Users are unhappy with trend indicators and want actual estimates.
- Spend estimates are too noisy to be useful.
- Users treat spend estimates as definitive and are surprised when the estimate changes when the RTEs are released.
- Users compare these to other spend estimates and find discrepancies.

These risks are related to the accuracy, coverage, representativeness, coherence and interpretability of the data. The preceding sections indicated that all ADS suffered quality issues in these areas. Recommendation 6 also highlighted that many users of RTI had difficulty using them correctly. Significant training may be required. Included in that would be making clear the understanding of the definition of tourism (recommendation 4) which in NZ’s case varied for RTIs from the standard definition. The need for training and education in the use of ADS should not be underestimated. Indeed clear training is likely the key work to make an ADS mainstream.

Recommendations 2 and 3 focused on the cleanliness of the data and the rules and mitigation strategies MarketView used to address these issues. The new ‘Aspiring’ data is a significant value add in understanding the nature of the merchants, their business and location. This sort of value
add is critical to high quality transaction data and only comes through long term use of a dataset. There is a high risk that ADS derived data will have significant errors until there is an investment in data cleaning and augmentation.

NZ had trouble with coherence of the data, (i.e. being able to make a comparable time series). Another related study (NZ 2012) in the use of credit cards by international visitors found that visitors from different countries have very different behaviour in the use of foreign cards. This makes it challenging to compare foreign tourism except on a per country basis. Even then the use of credit cards compared to cash changes over time and will vary by source country as well.

The significant changes in the recommendations highlights the need to continue with a reliable indicator of tourism against which ADS are tested. It also highlights the value in widespread use of a product to identify its value and issues.

Different types of expenditure have a different capacity (and infrastructure) for prepayment. Tracking prepayment is challenging. In general ADS data is better if it can be linked. If online travel agents, telecommunication location services and electronic transactions can be linked then a much clearer picture is possible. The nature of change in the service industry means that new ways of purchasing services become available regularly (such as Airbnb). ADS will need to constantly change to include these new forms. Dealing with the pace of change is a real risk for ADS.

In the European Union (EU), data access and use will be governed by the General Data Protection Regulation. This imposes strict requirements that data can only be used in accordance with the purposes that users signed up for when data was collected. It also restricts data transfer outside the EU. These are not too dissimilar to the Australian Privacy Act which says that data should be used for disclosed primary purposes, except the privacy act also allows reasonable secondary purposes that stem from the primary ones. As ADS deal with personal information there is the risk of legislative change affecting what can be collected.
4 Roadmap for inclusion of alternative data sources

Almost all tourism data consumers see great value in the current mainstream data sources. They do not want to see those disappear. Government action, particularly around reports and planning requires accurate data. This report has discussed the drivers for ADS: the need to make decisions at finer spatial and temporal scales and the need to understand the impact of data (reporting, business, marketing, activity), particularly the impact of marketing with the immediacy that the market demands.

Accuracy was the first requirement almost all tourism data consumers mentioned. Yet accuracy is highly dependent on other quality criteria and these quality criteria are yet to be fully explored for ADS.

There is a temptation to blend ADS and mainstream data products into one set of tourism statistics. This is technically possible however there are social and institutional challenges. First, the best insights are available when the data products can be linked at unit record levels. Commercial providers of ADS may not have sufficient consent from the individuals represented to support this. There may also be commercial reasons not to share unit records. Second, the quality of ADS is untested and, more importantly, unknown to tourism data consumers. Without a good understanding of the quality constraints of mainstream data products and ADS then it is unlikely tourism data consumers will be able to appreciate the quality constraints of the blended product. Third, if ADS were blended into mainstream data products it is likely that there would be continued calls for the unblended ADS product. The blended ADS becomes invisible to tourism data consumers and therefore the unblended product becomes something to still aspire to.

In the light of this, Data61 is describing a way forward with two main themes to its roadmap:

- Develop a set of explicitly labelled ADS based tourism services. These are to be tested and evaluated over a five year term.
- Explore ways of funding the ADS tourism services.

Each of these will be considered in the sections below.

4.1 ADS based services

Some tourism data consumers are already exploring ADS. In the short term this is good as it drives innovation. Formal work with ADS could end there. Tourism data consumers could purchase ADS as required from industry. The private sector would eventually settle on tourism products that fit the market. There is some value in this as the data provision industry improves data quality and data cleaning techniques and the tourism data consumers work through what is important to them. On the downside a tourism data consumer led development of ADS might create inconsistency of data across consumers and potentially reduce the public transparency of actions based on tourism data.

Alternatively, the variety of approaches pursued now, can advise a consistent long term approach. Data61 proposes that the tourism community work towards a clearly labelled set of ADS based
tourism services. The services would be clearly labelled as online to reflect their digital source and distinguish them from mainstream products. The tourism data types covered by the services would depend on the data sources involved. For example, an Online Tourism Transaction Spend and Movement Service (because of the high and medium rating) would return data on spend and location. That is, not just spend but some dispersal information also. The distinct name will help deal with some of the risks in comparing the new services with mainstream data products.

The exact products would come out of a workshop program involving tourism data consumers, custodians of ADS (banks, telecommunications companies, OTA, social media representatives) and data infrastructure providers (data linkage vendors). The analysis of ADS (Table 2) indicated that each ADS class had different strengths. The workshops would explore combining ADS in the same class to improve quality and linking ADS across classes to improve the utility across more tourism data types. This came out in the Spanish study (section 2.3.3) where telecommunications data was linked with transaction data to provide good information on spend and movement.

These services could be as simple as a Regional Spend Estimate (RSE) (section 2.3.1). A RSE could provide spend across select spatial areas and over select time periods. If enough ADS are linked there could be sufficient complexity to support a TableBuilder7 like interface where arbitrary aggregations can be retrieved. Another alternative is a service that supports privacy preserving operations. That is, statistical operations (such as analysis of regressions and means) over unit record data that do not reveal personal information.

As these datasets are primarily from the private sector, the private sector will play a key role. The funding model (next section) might determine if the private sector competes on the same services (NAB online tourism transaction spend and movement estimate vs the Westpac online tourism transaction spend and movement estimate) or whether consortiums form behind specific services. Secure linking platforms can be used to bring together data from multiple sources in a secure environment preserving privacy. Such platforms could also combine the tourism data that tourism data consumers collect from their own contacts. What is required is that data custodians have privacy policies which support this and get consent from the individuals.

The employment sector has similar services. Private sector companies such as Burning Glass Technologies, provide employment data sources alongside the mainstream sources. Then other bodies, such as the Australian Government’s Internet Vacancy Index (IVI8), publish an index based on these sources clearly labelled as ‘internet’ based.

The workshops provide an opportunity for tourism data consumers to articulate the product they want, at the spatial and temporal scale they require, using the best sources and technology that can be mustered. For example, while the data may exist for use by the data custodian, under what terms and conditions, price, and so on can the data be made available?

The choice of data source will also determine the capacity of the product to track international travellers. Many ADS are most effective with national travellers. The emphasis on international travellers will need to be weighed when designing the service.

With the tourism data services deployed and accessible to tourism data consumers the quality of the services can be evaluated over time and the utility of the statistics for the purposes of the tourism data consumers can be determined. Formal case studies for review will need to be established. Building trust in what these services can deliver will take time. The suitability of the various ADS to inform the tourism data features presented in section 3.3 has been informed by a qualitative assessment. This needs to be verified through quantitative means where possible, investigating for each ADS class the "fitness for the proposed purpose".

A suitable timeline for the ADS services might be:

- Year 1: Stakeholders agree on the first round of ADS services. Stakeholders include tourism data consumers, ADS custodians and linkage infrastructure providers.
- Year 1: A funding model is determined.
- Year 2-3: Interest sought from the public and private sector to build the agreed ADS data services. If ADS from more than one ADS custodian are chosen to form an ADS service there will be a period of negotiation between the ADS custodians. The service provider may also need to spend time cleaning the data before deploying the service.
- Year 3-4+: Annual workshops on use of ADS services as well as an analysis of their quality. A priority of the workshops would be to educate tourism data consumers about the appropriate use of the ADS services.
- Year 5: A review of the ADS services to evaluate their quality and address any issues.

This could repeat, possibly in parallel for further ADS services as required.

Apart from privacy concerns the private sector may have commercial reasons not to be involved. These ADS services would require nationally agreed standards on how to measure tourism. Standards may work against a business model which prioritises working with the customer to achieve the customer’s outcome. Tailored analytics is a high value service, standardised data products can drive costs down as processes become routine and increase competition to produce comparable products.

By moving the use of ADS into a common space accessible to tourism data consumers there will be greater sharing of insight and a clearer view as to the role ADS play. If the PC’s Data Availability and Use recommendations are adopted then there is an option of nationalising private data in the national interest. Perhaps this possibility might spur the private sector to increase the amount of privacy preserving linkage across ADS.

---

Figure 8 Tourism data roadmap

Apart from privacy concerns the private sector may have commercial reasons not to be involved. These ADS services would require nationally agreed standards on how to measure tourism. Standards may work against a business model which prioritises working with the customer to achieve the customer’s outcome. Tailored analytics is a high value service, standardised data products can drive costs down as processes become routine and increase competition to produce comparable products.

By moving the use of ADS into a common space accessible to tourism data consumers there will be greater sharing of insight and a clearer view as to the role ADS play. If the PC’s Data Availability and Use recommendations are adopted then there is an option of nationalising private data in the national interest. Perhaps this possibility might spur the private sector to increase the amount of privacy preserving linkage across ADS.
4.2 Funding ADS services

There was little support for state agencies to pay more for tourism data. This report is not an in-depth cost-benefit analysis of options, however Data61 have listed some obvious funding options.

4.2.1 User pays

While states contribute to funding some of the cost of the IVS and NVS, many tourism data consumers pay very little for tourism data at this stage. As the benefits are received by the tourism data consumers they could be requested to pay for these additional services. This is particularly true for the marketing purpose which is a key driver for the finer spatial and temporal scales.

As ADS are primarily from the private sector it is reasonable to suggest a commercial solution to funding. The role of the current trusted mainstream data providers would then be to develop the national minimum quality parameters of ADS services and standardise the data products where appropriate.

4.2.2 Publicly funded

Request the various levels of government fund the delivery of the services (otherwise similar to user pays). The government would need to prioritise whose data needs would be satisfied first.

4.2.3 Reducing the cost of mainstream tourism data

To fund the use of ADS, the cost of mainstream data could be reduced. As mentioned in section 4.1 ADS are more readily suited to the monitoring of national travel. Therefore, the likely cost reductions are through national mainstream data products rather than international products. TRA have provided details of at least 80 overseas countries which measure international visitation through surveys due to the complexities of defining and collecting tourism activity on this type of traveller. It is the primary collection technique for international travel data worldwide.

The NZ experience suggests that the surveys offer many features that cannot be replaced with ADS. With the quality of ADS still unclear, the quality of the decisions based on ADS would come into question. Stopping the surveys now and switching exclusively to ADS would be high risk in Data61’s opinion. The proposed ADS services program covers multiple years. It seems reasonable to wait until the program has developed quality services and the level of trust in those services is improved. Continuing the surveys allows quality comparisons and evaluation of alternative sources of calibration.

Some options for reducing the cost of the NVS surveys were discussed with Phil Hughes from ORC (2017a) who developed the methodology behind the IVS and NVS surveys. All these approaches could save money but all come at a statistical cost, i.e. a cost in quality of those products. A cost benefit analysis has not been done.

TRA has been investigating geolocating the home base of phone numbers (mobile and landline). A spatially targeted phone survey looks promising to reduce the number of calls required to gain improved spatial distribution of the sample, therefore improving accuracy and reducing costs.
It may also allow targeting of areas of relatively low population where there is evidence in the past of primarily short range tourism. When properly compensated for this would increase the quality of the data at finer spatial scales.

It is also possible to use online surveys rather than a personal interview. Szolnoki and Hoffmann (2013) have reported that online surveys can be effective but only if the representativeness of the sample is controlled. Uncontrolled online surveys produce highly skewed, unrepresentative results which are difficult to compensate for. Skarupova (2014) suggested the additional care for online survey preparation and effort in compensating for unrepresentative results may severely erode any cost savings. A low cost online solution is a panel survey (NSS 2017) however analysing the quality cost and benefit of a panel survey is outside the scope of this report.

While an ADS is being established is not the time to be cutting costs on mainstream products. Instead the continuing high quality mainstream products need to be compared with the ADS products. Some of the ADS will have historical information for immediate comparison. However the best case ADS services will see a combination of ADS classes and a combination of ADS within a class all linked. This linked data product will need time to develop a time series to gauge its coherence.

The ADS service timeline reaches to five years and then a review. By that stage it would have been operational for only two to three years. It seems likely (as happened in NZ) that the review may result in a change of methodology, requiring the time series to be re-established and require further testing against mainstream data products.

### 4.3 Changing the survey operation

Data61 would like to cover two more scenarios which suggest significant changes for surveys: artificial intelligence and person-centric ADS capture. There are glimpses that these might be possible over the next 10-20 years. With the cost of the survey products pegged to sample size and the sample size increasing as the population increases, the survey cost will continue to grow unless changes are made.

#### 4.3.1 Artificial Intelligence (AI)

AI is entering a stage where it is becoming mainstream. Digital assistants and chat bots are commonplace. The NVS is conducted by computer assisted phone interview. The interviewer develops a rapport with the interviewee to hold their attention during the survey and draw out the desired information.

This is currently beyond the capabilities of AI. It requires AI to pass the Turing test, a benchmark for AI that is still to be accomplished. In a Turing test messages are passed between a person and a responder. An AI system passes the Turing test if, over multiple engagements, a person cannot distinguish between a human and an AI responder.

Interacting with people is complex. This field will continue to grow and is one to watch. It is unlikely however that IVS could be replaced with AI due to the complexity of these person to person interactions and the need to efficiently identify international visitors with particular demographic characteristics prior to interview.
4.3.2 Person-centric ADS capture

During the Productivity Commissions (PCs) Data Availability and Use inquiry there were some submissions calling for a national database of financial transactions. This was seen as a real boon for researchers. The PC did not pursue this. Instead they proposed a comprehensive right to access, correct and transfer data. One of the use cases envisaged was a way of analysing personal data based on recruiting a sample of individuals and obtaining their consent and instruction to transfer their consumer (transaction) data to the researcher.

In the case of tourism this is more complex as an individual is likely to have used multiple merchants (online travel agents, finance providers, telecommunication companies) in the process of their travel. Each of these would need to be asked to transfer consumer information to the researcher. There are personal data aggregator systems that can allow a person to pull together the threads of all their personal data from different merchants. It is possible that these could be used to make it easy for an individual to give consent and instruction to provide data to a researcher. Incentives from travel providers could be given to participate. Apple already provides assistance for researchers who want to collect fitness information with its ResearchKit⁹.

Individuals would also be prompted with a short, online survey, when the ADS data indicated that the person had travelled, to fill in any missing information. It effectively sets up a panel survey where most of the questions are answered by data transfer rather than personal recall. Keeping the personal impost low may encourage retention on the panel for extended periods and allow for a large panel size.

The right to transfer data is only a recommendation of the PC, the government will provide a response in early 2018. It also does not cover international visitors. It does highlight that it is easier to know about the travel of Australians than internationals. This is as true overseas as in Australia. If this approach could also capture tourism data for Australians travelling overseas then perhaps a de-identified aggregate database of Australian travel overseas could be traded for a similar database of international visitors coming to Australia from specific countries.

5 Conclusion

Data is everywhere. Tourism data consumers have been effectively using mainstream data products for reporting, business, marketing and activity. With so much data around, expectations are raised in the ability to carry out analyses at more and more detailed spatial and temporal scales. Users expect to have data shortly after collection with the same quality they have experienced at the national and state level. Unfortunately the mainstream products are not designed to deliver the same quality at finer scales which is causing friction in the tourism data community, especially for marketing purposes.

The abundance of data in ADS looks promising. However, in practice, privacy and commercial concerns mean that current practical examples of ADS based data products operate over a much smaller pool of data, often restricted to a single ADS data custodian such as a bank. These restricted ADS services are likely to be challenged across a range of quality criteria. Data61 has recommended tourism stakeholders (users, ADS custodians and systems experts) get together to try and agree to combine and link many ADS before deploying them as ADS services. These services will still have a distinctly different quality to mainstream services and so must be distinctly named and followed up with education and training.

Through making these services available to the tourism community, that community (including their current trusted providers of mainstream data products) can examine the issues with ADS data and determine the right use for that data. Tourism data consumers do not want to throw out mainstream data products without proven replacements.

ADS services will involve the private sector and the funding mechanism to create them must be determined in the tourism community. While ADS may, in the long term, reduce some of the data capture needs in mainstream tourism products, doing so in the short term will make it hard to compare the quality and utility of ADS. To become mainstream data products ADS need to find their accepted niche, what they do best and more cost effectively than existing mainstream products.
Glossary

**Accuracy** (ABS): indicators should have a degree of precision to which the indicator correctly describes the phenomenon being measured.

**Alternative Data Sources**: Data sources that reveal information about tourism, that have not traditionally been used in measuring tourism activity. Examples include: data from mobile phones, credit cards, social media, and reservation systems.

**Coherence** (ABS): indicators should be comparable across industries and presented as a time series. The latter is most important to tourism data consumers.

**Coverage** (ORC): indicators should include all in-scope tourism activity. For example all forms of expenditure (electronic, cash, prepayment) in a spend indicator.

**Institutional environment** (ABS): indicators should be selected from credible sources. Sources need to use consistent methodology and accessible methodology.

**Interpretability** (ABS): indicators should be accompanied with supporting information that can be understood by the general reader.

**Latency or Timeliness** (ABS): indicators should have minimal delay between the reference period and the release of the data.

**Linkability or single source** (ORC): indicators should be either available from a single source or linkable so that relationships among variables can be examined.

**Mainstream Data Sources**: International Visitor Survey (IVS), National Visitor Survey (NVS), Australian Bureau of Statistics data, and Bureau of Infrastructure, Transport, and Regional Economic data.

**Representativeness** (ORC) indicators should provide coverage of all the components of tourism activity so that no component is over- or under-represented. For example all age segments of a population should be represented in the data. Sensitivity relates to the indicator being able to respond when the tourism activity changes. An unrepresentative indicator will not be sensitive to changes in those components not represented.

**Scalable** (ABS): indicators should be amenable to aggregation or disaggregation by topics of interest, e.g. industry, product, geography and time scales.
### Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABS</td>
<td>Australian Bureau of Statistics</td>
</tr>
<tr>
<td>ADS</td>
<td>Alternative Data Sources</td>
</tr>
<tr>
<td>AHA</td>
<td>Australian Hotel Association</td>
</tr>
<tr>
<td>AIFS</td>
<td>Australian Institute of Family Studies</td>
</tr>
<tr>
<td>AIHW</td>
<td>Australian Institute of Health and Welfare</td>
</tr>
<tr>
<td>ARA</td>
<td>Accredited Release Authority</td>
</tr>
<tr>
<td>BITRE</td>
<td>Bureau of Infrastructure, Transport, and Regional Economics</td>
</tr>
<tr>
<td>BNZ</td>
<td>Bank New Zealand</td>
</tr>
<tr>
<td>GDS</td>
<td>Global Distribution System</td>
</tr>
<tr>
<td>GDP</td>
<td>Gross Domestic Product</td>
</tr>
<tr>
<td>GVA</td>
<td>Gross Value Added</td>
</tr>
<tr>
<td>IVS</td>
<td>International Visitor Survey - Australia</td>
</tr>
<tr>
<td>LGA</td>
<td>Local Government Area</td>
</tr>
<tr>
<td>NAB</td>
<td>National Australia Bank</td>
</tr>
<tr>
<td>NDC</td>
<td>National Data Custodian</td>
</tr>
<tr>
<td>NID</td>
<td>National Interest Dataset</td>
</tr>
<tr>
<td>NVS</td>
<td>National Visitor Survey - Australia</td>
</tr>
<tr>
<td>NZ</td>
<td>New Zealand</td>
</tr>
<tr>
<td>OAD</td>
<td>Overseas Arrivals and Departures</td>
</tr>
<tr>
<td>ORC</td>
<td>ORC International</td>
</tr>
<tr>
<td>OTA</td>
<td>Online Travel Agents</td>
</tr>
<tr>
<td>PC</td>
<td>Productivity Commission</td>
</tr>
<tr>
<td>PRTE</td>
<td>Provisional Regional Tourism Estimate</td>
</tr>
<tr>
<td>RTE</td>
<td>Regional Tourism Estimate</td>
</tr>
<tr>
<td>STA</td>
<td>Survey of Tourist Accommodation</td>
</tr>
<tr>
<td>STO</td>
<td>State and Territory Tourism Organisations</td>
</tr>
<tr>
<td>TRA</td>
<td>Tourism Research Australia</td>
</tr>
</tbody>
</table>
References


ORC International (2017) *Can big data replace the Australian tourism surveys?* Internal discussion paper from ORC.


Productivity Commission (2017), Data Availability and Use, Report No. 82, Canberra.


Sanaz Shafiee and Ali Rajabzadeh Ghatari (2016) Big Data in tourism industry. 2016 10th International Conference on e-Commerce in Developing Countries: with focus on e-Tourism (ECDC), Isfahan, 2016, pp. 1-7


AT CSIRO WE SHAPE THE FUTURE
We do this by using science and technology to solve real issues. Our research makes a difference to industry, people and the planet.

FOR FURTHER INFORMATION
Rod Battye
A/g Assistant General Manager
Tourism Research Australia
t +61 2 6272 6879
e rod.battye@tra.gov.au
w www.austrade.gov.au

Gavin Walker
Principal Experimental Scientist | Market Design
t +61 2 6216 7030
e Gavin.Walker@csiro.au
w www.data61.csiro.au