A Bimodal Accessibility Analysis of Australia Using Web-based Resources

Sarah Meire[1]
Ben Derudder[2]

Abstract: *A range of potentially disruptive changes to research strategies have been taking root in the field of transport research. Many of these relate to the emergence of data sources and travel applications reshaping how we conduct accessibility analyses. This paper, based on Meire et al. (in press) and Meire and Derudder (under review), aims to explore the potential of some of these data sources by focusing on a concrete example: we introduce a framework for (road and air) transport data extraction and processing using publicly available web-based resources that can be accessed via web Application Programming Interfaces (APIs), illustrated by a case study evaluating the combined land- and airside accessibility of Australia at the level of statistical units. Given that car and air travel (or a combination thereof) are so dominant in the production of Australia's accessibility landscape, a systematic bimodal accessibility analysis based on the automated extraction of web-based data shows the practical value of our research framework. With regard to our case study, results show a largely-expected accessibility pattern centred on major agglomerations, supplemented by a number of idiosyncratic and perhaps less-expected geographical patterns. Beyond the lessons learned from our case study, we show some of the major strengths and limitations of web-based data accessed via web-APIs for transport related research topics.*

Keywords*:* "web-based data", "application programming interfaces (APIs)", "road and air transport", "bimodal accessibility", "Australia".

1. <u>Introduction</u>

A range of potentially disruptive changes to research strategies have been taking root in the field of transport research. Many of these relate to the emergence of new data sources and travel applications reshaping how we conduct accessibility analyses. Although 'big data' can be an inflated and hyped term (Ahmadi, Dileepan, & Wheatley, 2016; Jagadish, 2015), it is clear that new types of often-extensive datasets are increasingly supplementing more conventional data sources and approaches to data gathering such as activity-travel diaries and retrospective interviews (Tranos & Mack, 2018; Witlox, 2015). Examples of big data that can be used in transport and accessibility studies include, among others, mobile phone call detail records (see, e.g., Becker et al., 2011; Demissie, Correia, & Bento, 2015; González, Hidalgo, & Barabási, 2008; Kung, Greco, Sobolevsky, & Ratti, 2014), Bluetooth data (see, e.g., Barceló, Montero, Marqués, & Carmona, 2010; Hainen et al., 2011; Malinovskiy, Saunier, & Wang, 2012; Versichele et al., 2014), social media data (see, e.g., Mogaji & Erkan, 2019; Rashidi, Abbasi, Maghrebi, Hasan, & Waller, 2017; Serna & Gasparovic, 2018; Zhang & Zhou, 2018), as well as data acquired by means of global positioning systems (see, e.g., Cui et al., 2016; Stipancic, Miranda-Moreno, Labbe, & Saunier, 2017; Wong, Szeto, Wong, & Yang, 2014; Zuo, Wei, & Rohne, 2018). Against this backdrop, consumer-oriented travel data provided by online route planners, meta-search engines and/or web-crawling services (e.g. Cheapflights, Connections, Google Maps, etc.) provide new and potentially rich data sources in the field of transport research in general and accessibility research in particular. This paper, based on Meire et al. (in press) and Meire and Derudder (under review), aims to explore the potential of some of these

[1] Ghent University, Social and Economic Geography, Sarah.Meire@UGent.be
[2] Ghent University, Social and Economic Geography, Ben.Derudder@UGent.be

data sources by focusing on a concrete example: we present a bimodal accessibility analysis of Australia based on publicly available, web-based resources that can be accessed via web Application Programming Interfaces (APIs). Australia is an interesting case to conduct bimodal accessibility analyses as it is characterised by a spatially dispersed settlement pattern: the distances between the main population centres are on average relatively large. Since public transport services such as train and bus are relatively unimportant to connect centres on the national scale, or often even almost completely absent outside the urban areas, (private) road and air transport are the only viable alternatives to cover often vast distances (Donehue & Baker, 2012; Nutley, 2003). As such, given that car and air travel (or a combination thereof) are so dominant in the production of Australia's accessibility landscape, a systematic bimodal accessibility analysis based on the automated extraction of web-based data shows the practical value of our research framework. The objective of this paper is therefore twofold: (1) to introduce a framework for transport data extraction and processing using publicly available web-based resources, and (2) to enhance our understanding of the uneven geographies of accessibility across Australia. To this end, we evaluate the accessibility of locations from a car-air-car travel perspective (or any subset of combinations as long as it produces the shortest travel time) using (shortest) travel time as the primary indicator of accessibility. Focusing on passenger transport, we incorporate and combine both road and air travel to quantify how fast people can travel from every statistical area to all other statistical areas. This combination of land- and airside accessibility consists of three route segments: (1) travel from the origin to a departure airport using the road network, (2) air travel (including transfer time in case of connections requiring a stopover), and (3) travel from an arrival airport to the destination using the road network. In addition to this, unimodal car travel is also taken into account. In this respect, our approach conceptually resembles the bimodal accessibility perspective in Redondi, Malighetti, & Paleari (2011).

2. Literature review

Previous studies adopting web-based data sources in transport research (e.g. Dumbliauskas, Grigonis, & Barauskas, 2017; Fuellhart, Derudder, O'Connor, & Zhang, 2015; Grubesic & Zook, 2007) reveal that there are three predominant approaches through which consumer-oriented, web-based travel data may be extracted, and which largely depend on the resources available.

A first approach consists of using a web API developed by a meta-search engine or online route planner. An API is an interface to communicate with a web service in order to be able to access its content. Through the implementation of an API in a programming script, large volumes of data can be efficiently extracted and processed. Examples of overland travel APIs include, among others, those provided through the Google Maps Platform that allow accessing Google's anonymised and aggregated travel data, collected from smartphone users (Dumbliauskas et al., 2017; García-Albertos, Picornell, Salas-Olmedo, & Gutiérrez, in press). Although Google's route planner has not yet been widely adopted in scientific research (García-Albertos et al., n.d.), a number of research projects have used Google's travel API services. To the best of our knowledge, Gu, Wang, & McGregor (2010) were the first to adopt the Google Maps API in their study on the optimization of preventive health care facility locations. In Padeiro (2018), the Google Maps Directions API is used to evaluate the pedestrian access of elderly people in metropolitan Lisbon to community pharmacies. The Directions API provided by the Google Maps Platform is adopted by García-Albertos et al. (in press), Wang & Xu (2011) and Xia et al. (2018). In Wang & Xu (2011), for example, a desktop tool is developed that implements the Google Maps Directions API in order to automatically estimate an origin-destination travel time matrix, after which the results are compared with travel times generated through the use of the ArcGIS Network Analyst module. Xia et al. (2018), in turn, propose an accessibility framework that integrates both travel costs and potential opportunities at a variety of scales in Australia. Each of the abovementioned studies adopts a static perspective on accessibility, and therefore does not account for changing travel times during the day due to variations in congestion levels and/or destination attractiveness. Yet more dynamic uses are also possible, as shown by the use of the Google Maps Directions API by García-Albertos et al. (in press) in a longitudinal analysis of urban accessibility in Madrid. In addition to the Google Maps Directions API, the Google Maps Platform also provides travel information through its Google Maps Distance Matrix API service. Dumbliauskas et al. (2017), for example, use the Google Maps Distance Matrix API complemented with open source software in order to conduct a unimodal travel time analysis of Kaunas, Lithuania.

In addition to the different APIs provided by Google, there are a range of other travel APIs. In a study by Hajinasab, Davidsson, Holmgren, & Persson (2017) on the use of online services for transport simulation models, the API of the local online travel planner Skånetrafiken is adopted to collect data on available public transport options. This API is complemented with the Google Maps Directions API to gather cycling, walking and driving routes, together with a weather forecasting service API. Niu, Wang, Xia, Wu, & Tang (2018), in turn, use an API provided by the Chinese local map service provider AMap to gather shortest travel durations between a range of origin points and park entrances in Wuhan, China. Based on these data, they evaluate the accessibility and effective service ratio of Wuhan’s main urban parks. A comparative analysis of the Google Maps API, the Bing Maps API and the MapQuest API was conducted by Socharoentum & Karimi (2015), showing that the different APIs might generate slightly different results.

Such an API-based approach can of course not be adopted if there is no API available or in case the API does not meet the concrete needs of the research question at hand. A second, alternative approach therefore consists of ‘screen-scraping’ online route planners, meta-search engines and/or web-crawling services using programming code and/or specific software packages (e.g. ParseHub), albeit that often a number of legal constraints must be considered in such cases. In such an approach, web pages’ content is ‘scraped’ in order to automatically extract the travel options they display. To the best of our knowledge, the only studies applying web-scraping to gather travel data were conducted by Grubesic, Horner, Zook, & Leinbach (2006) and Grubesic & Zook (2007). Through the use of a Perl script and several object-oriented modules therein, they automate online flight searches on Expedia.com to collect flight ticket options from a Global Distribution System. As such, the collected data represent (real-time) ‘*choice-based information provided to consumers during the booking process*’ (Grubesic et al., 2006). Grubesic & Zook (2007) then use these data to analyse air travel accessibility in the United States in terms of flight connections, flight time and ticket costs.

A third and final approach consists of manually browsing the meta-search engine, web-crawling service or online route planner to gather air/car travel data, although this obviously entails a more time consuming process in case an excessive dataset is required. In Fuellhart et al. (2015), for example, the travel search engine Skyscanner is used to manually access the lowest available airfares for 226 inter-city connections throughout Australia. Hence, paralleling the approach adopted in Grubesic et al. (2006) and Grubesic & Zook (2007), they apply a ‘consumer perspective’ to accessibility by mimicking the consumer’s online flight search before the actual booking takes place.

In this paper, we will focus on web-based data accessed via two specific web-APIs for transport related research topics: Google’s QPX Express API and the Google Maps Distance Matrix API.

3. Study area, data and method

3.1. Study area

The analysis was carried out at the Australian Statistical Areas Level 2 (Australian Bureau of Statistics, 2016b), including 2289 out of 2310 statistical areas. Twenty-one statistical areas were thus excluded from the analysis, involving eighteen non-spatial statistical areas and three statistical areas consisting of an island with neither a bridge to the Australian mainland nor an airport. The population weighted centroids of the statistical areas, modelled in ArcGIS using a 1x1 km population grid in raster format (Australian Bureau of Statistics, 2016a), served as the points of origin and destination. However, due to insufficient population data and the spatial configuration of some statistical areas, 57 population weighted centroids were replaced by geometric centroids using ArcGIS. These 57 geometric centroids were near-randomly distributed within Australia (see Figure 1). During the data collection process, 21 centroids (of which 18 population weighted centroids and three geometric centroids) had to be manually and marginally moved towards the road network in order to rectify geocoding errors. With respect to the airside accessibility, 159 Australian airports were included in the analysis. These airports are a subset of the 317 certified and/or registered airports providing regular public transport services or having (potential) charter use (Australian Airports Association, 2012), since we only included the airports that are commercially accessible as evidenced by their presence in meta-search engines and/or web-crawling services (i.e. Skyscanner, Google Flights and/or ITA Matrix). In this way, we take a consumer's perspective throughout the data collection process by mimicking the booking process of travellers. The study area is visualised in Figure 1.

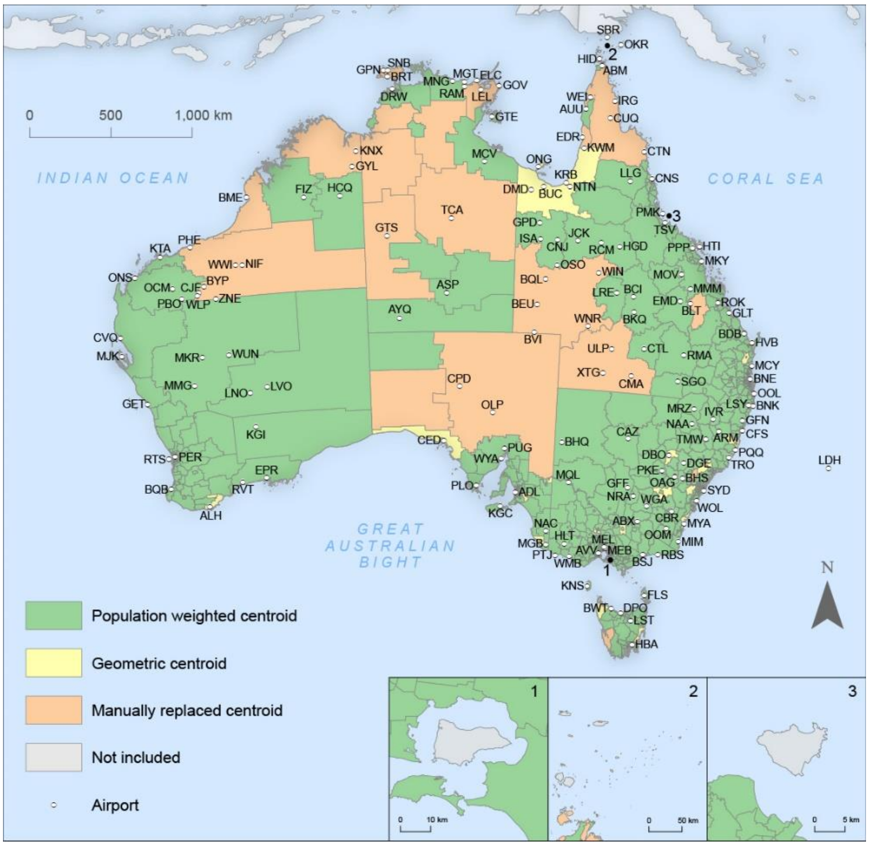

Figure 1 : Study area. Source vector data on SA2 boundaries are obtained from the Australian Bureau of Statistics (2016b). Source vector data on country boundaries are obtained from Esri (2018). (Meire and Derudder, under review)

3.2. <u>Data</u>

In order to map the combined land- and airside accessibility within Australia in terms of travel time, road and air travel data were collected.

With regard to air travel, we collected web-based data on scheduled flight itineraries through the use of a publicly available travel API. Hence in parallel to Fuellhart et al. (2015), Grubesic et al. (2006) and Grubesic & Zook (2007), we mimic the online flight search of travellers before the actual booking takes place. To this end, Google's web-based QPX Express API was implemented in a Python 3.6 script in order to automatically gather real-time data on flight durations between each pair of airports. As most air transport databases only provide data on individual flight legs (Derudder & Witlox, 2005b, 2005a), Google's QPX Express API allows the user to collect data on full, scheduled flight itineraries (origin-destination data). The API is based on QPX Software developed by ITA Software, which '*uses algorithms to combine and parse multiple sets of flight information from airlines, including pricing and availability data, to create an up-to-date database that can be searched across*' (Google Company, 2018a, 2018b). With the aim of reducing the influence of booking time and seasonal fluctuations, scheduled flight data were collected for three different departure dates (i.e. Monday 16 April 2018, Thursday 16 August 2018 and Sunday 16 December 2018), after which the median value (of the fastest flights) was used in subsequent calculations in order to mitigate possible outliers. The air travel data acquisition took place on 13 February for the first departure date, on 14 February 2018 for the second departure date, and on 15 and 16 February for the third departure date. Since 159 airports are included in the analysis, 75 366 API requests were sent (25 122 for each departure date). We provided Google's QPX Express API's request body with the following input data: one adult passenger, the departure and arrival airport, the date of departure, a maximum of five transfers, no obligation of refundable fares and a request to return as many solutions as possible; supplemented by the API key, the URL and the headers. After each flight request, Google's QPX Express API generated a response body containing the air travel data in a JavaScript Object Notation (JSON) format, a structured and lightweight data interchange format that is independent of a specific programming language (Bassett, 2015). These response bodies were saved in a shelf file using the Python 3.6 shelve module, allowing us to store the extracted data objects on the hard drive in order to be re-opened and retrieved in subsequent PyDev modules (Sweigert, 2015). The architecture of a shelf file follows a dictionary-like format, meaning a key or object ID (in our case, the airport pair) is linked to a certain value (in our case, Google's QPX Express API's response body). We then extracted the shortest flight duration per airport pair and per date from the JSON response bodies, after which the median shortest air travel time between each pair of airports was quantified.
In order to empirically evaluate the accuracy of Google's QPX Express API's data on flight durations, we compared the shortest flight durations generated via Google's QPX Express API with data extracted from Skyscanner (www.skyscanner.com), a well-known (travel) web-crawling service. To this end two random samples of 50 origin-destination airport pairs were selected for each departure date using a Python script, for which the shortest flight durations were extracted via both data sources. The two samples respectively represent (1) airport pairs between which Google's QPX Express API could not generate any flights, and (2) airport pairs between which Google's QPX Express API generated at least one flight itinerary. As such, 300 out of 75 366 queries were manually conducted via Skyscanner (on 20 and 21 February 2018). For 141 out of 300 flight queries, however, neither data source could generate a travel option. In four cases Skyscanner generated a flight itinerary while Google's QPX Express API did not. In contrast, in twelve cases Google's QPX Express API generated a flight itinerary while Skyscanner could not find any flights. The shortest travel time results generated by Google's QPX Express API and Skyscanner were then plotted against each other (Figure 2). Although a number of outliers are observed, we may assume that both meta-search engines generate similar (shortest) flight durations.

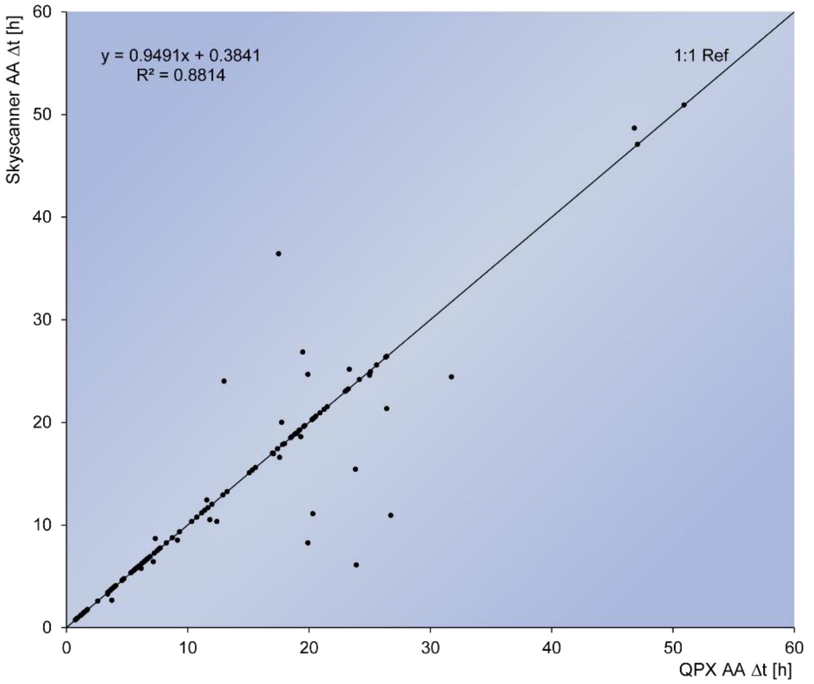

Figure 2 : Comparison between Google's QPX Express API and Skyscanner data. The X-axis represents the shortest flight duration between random airport pairs (on 16 April 2018, 16 August 2018 and 16 December 2018) generated via Google's QPX Express API. The Y-axis represents the shortest flight duration between the same airport pairs generated via Skyscanner.

With regard to the landside accessibility, three sub-components can be identified: (1) the overland travel time between the point of origin and the departure airport, (2) the overland travel time between the arrival airport and the final point of destination, and (3) the overland travel time between the origin and destination centroids in case they are geographically so close that including an air travel segment would not improve travel time. The potential departure/arrival airports of the origin/destination centroids were selected based on a Euclidian distance criterion, which itself depended on Australia's Remoteness Area Structure (Australian Bureau of Statistics, 2011): if a statistical area's centroid is situated in an area that is considered to be 'Remote' or 'Very Remote' in terms of its relative access to services as outlined in Australia's Remoteness Area Structure, all airports within a Euclidian distance of 750 km from the centroid involved are considered to be potential departure/arrival airports. A 500 km distance limit was applied to all other centroids (i.e. those situated in 'Outer Regional Australia', 'Inner Regional Australia' and Australia's 'Major Cities'). These large distance limits were (arbitrarily) selected to ensure the inclusion of all potential departure/arrival airports while maintaining the feasibility of the overland data collection process. With regard to unimodal travel, all trips between centroid pairs situated within a Euclidian distance of 500 km from each other are considered potential unimodal (car) travel trips. In total, 41 091 origin centroid – departure airport pairs, 41 091 arrival airport – destination centroid pairs and 1 179 117 origin centroid – destination centroid pairs were identified.

Following the (primary) aim of this study, data on overland (car) travel times were collected by invoking the web-based Google Maps Distance Matrix API in a Python script. Since we did not specify a departure date nor time for the car travel component, no specific or real-time traffic/road conditions were taken into account, and we thus generated general values. The overland data acquisition concerning the origin centroid – departure airport pairs and the arrival airport – destination centroid pairs took place on 8 and 9 March 2018, respectively. From 10 March 2018 until 23 March 2018, the Python script was run for all origin centroid –

destination centroid pairs. The travel data extraction was executed on a 100 000 queries/day basis and the results were saved in Python shelf files.

3.3. Method

As previously mentioned, the combination of land- and airside accessibility consists of three route segments: (1) travel from the origin centroid to a departure airport using the road network, (2) air travel (including transfer time in case of connections requiring a stopover), and (3) travel from an arrival airport to the destination centroid using the road network. Depending on the centroid's position in Australia's Remoteness Area Structure, all airports within a Euclidian distance limit of 500/750 km from the centroid involved are considered to be potential departure or arrival airports. In case the origin and destination centroids are within a Euclidian distance of 500 km from each other, unimodal car travel is also taken into account. The land- and airside accessibility framework is visualised in Figure 3.

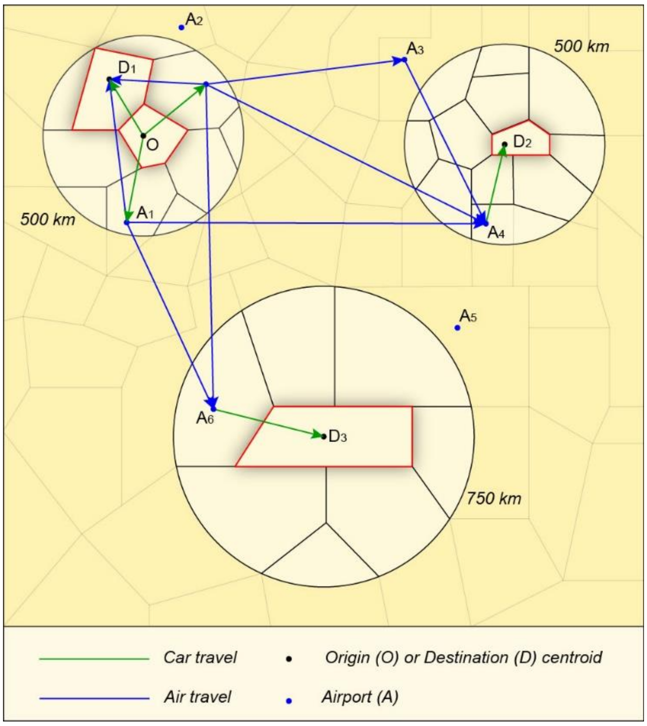

Figure 3 : Land- and airside accessibility framework

Using Python 3.6 software, we combined all possible route configurations with the aim of finding the shortest possible travel time between every pair of centroids. This combination process is visualised in Figure 4. The data collection process resulted in four shelf files, containing: (1) the overland, unimodal travel time between centroids that are within a Euclidian distance of 500 km from each other (the C_OD_O shelf), (2) the overland travel time between the origin centroids and their potential departure airports (the C_OA_D shelf), (3) the overland travel time between the potential arrival airports and the destination centroids (the A_AC_D shelf), and (4) the fastest flights (including transfer time if relevant) between every pair of Australian airports (the A_DA_A shelf). In order to determine the shortest travel time between a pair of centroids, we first search the C_OC_D shelf for the centroid pair involved. If the centroid pair is present, the corresponding unimodal travel time is temporarily considered to be the shortest travel time. If not, an infinite travel time is assigned to the centroid pair involved. In a next step, we search for all centroid-airport pairs in the C_OA_D shelf containing the origin centroid involved, and add these pairs to a new, empty Python dictionary. The same procedure is followed with regard to the airport-centroid pairs in the A_AC_D shelf, albeit that we search for airport-centroid pairs involving the destination centroid involved instead of the origin centroid. Thereafter, all centroid-airport pairs containing the origin centroid and all airport-centroid pairs containing the destination centroid (i.e. those present in the abovementioned Python dictionaries) are combined. We then examine each combination for an available airside segment. In case an airside segment is present in the A_DA_A shelf, the corresponding land- and airside travel times are aggregated. When the aggregated travel time is less that the previously defined travel time, the latter is replaced by the former. This procedure is repeated for all possible route configurations involving the origin and destination centroids involved in order to find the shortest travel time between them. Any remaining infinite travel times were replaced by the observed (shortest) maximum travel time. Finally, the mean shortest travel time for each centroid to reach all other centroids was calculated.

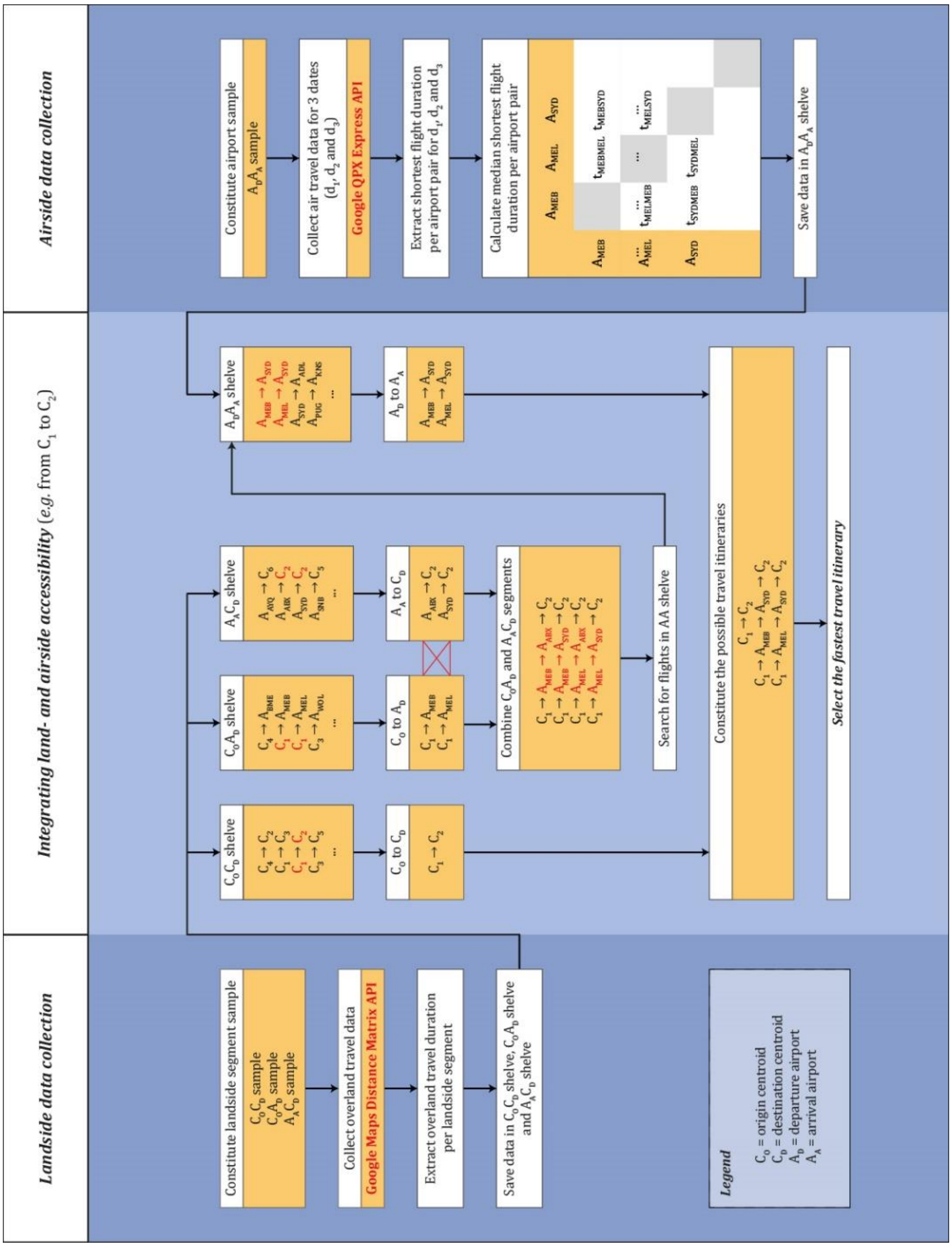

Figure 4 : Data collection and processing framework. BME, MEB, MEL, WOL, AYQ, ABX, SYD, SNB, ADL, PUG and KNS respectively correspond to Broome International Airport (Broome), Essendon Airport (Melbourne), Melbourne Airport (Melbourne), Illawarra Airport

(Wollongong), Connellan Airport (Uluru), Albury Airport (Albury), Kingsford Smith Airport (Sydney), Snake Bay Airport (Milikapiti), Adelaide International Airport (Adelaide), Port Augusta Airport (Port Augusta) and King Island Airport (King Island). The combinations of specific centroids and/or airports are for illustrative purposes only and therefore do not correspond to the actual combinations. (Meire and Derudder, under review)

4. Results

Figure 5 visualises the bimodal accessibility index of the Australian Statistical Areas Level 2. Results show that the southeastern part of Australia is generally characterised by high levels of accessibility, with a gradual transition from well accessible areas situated along Australia's coastline to less accessible areas situated more inland. To the (north)west of Sydney and to the east of Melbourne, however, a small number of less accessible statistical areas can also be observed, abruptly interrupting the abovementioned gradual transition. A similar abrupt low-accessibility zone embedded in a well-accessible matrix can also be observed to the west of Rockhampton, Queensland. Central and northern Australia are in turn for the greater part comprised of less accessible areas. Exceptions include the surrounding area of the Northern Territory's capital city Darwin and two centrally located statistical areas involving key sites of Australia's tourism industry in this region, i.e. Uluru and Alice Springs. In Western Australia, the most accessible areas are again situated along the coastline, whereas the less accessible areas are situated more inland or towards the north. Especially the region around Perth stands out as a high accessibility zone. Also the area near Karratha, one of Australia's small mining towns, is marked as a relatively high accessibility region.

Overall, the most populated cities are, unsurprisingly, hotspots of accessibility: the mean shortest travel time to reach all statistical areas is lowest in or nearby the most dominant cities. These cities are mainly located near Australia's coastline and are generally characterised by a hub airport. However, the results do not simply represent the configuration of Australia's airport system: the layout of the road network also plays a major role in rendering (in)accessible statistical areas. Access to main highways that are connected to relatively distant airports with a diverse and extensive network may lead to a higher accessibility overall. In South Australia, for example, the area around a number of low-service airports (i.e. Coober Pedy Airport, Ceduna Airport, Port Augusta Airport and Olympic Dam Airport) stand out as relatively accessible areas since they are situated alongside the state's main highways (i.e. the west-east directed Eyre Highway/Augusta Highway and the north-south direct Stuart Highway) which facilitate access to relatively distant and well-serviced airports – our mapping of this process conforms to what has been referred to in literature as 'air traveller leaking' (Ryerson & Kim, 2018).

The opposite pattern also emerges: the lack of main roads may prevent travellers from reaching well-connected (and sometimes even nearby) airports, consequently lowering the landside accessibility in particular and the overall accessibility in general. The abovementioned abrupt low-accessibility lobe to the northwest of Sydney, for example, is for the greater part comprised of wildlife area (i.e. the Wollemi and Blue Mountains National Parks), giving rise to a poor road network and, consequently, diminished accessibility. Similarly, the Alpine National Park lowers the accessibility index to the east of Melbourne to some extent. In the western part of Tasmania, Cradle Mountain-Lake St Clair and Franklin-Gordon Wild Rivers National Parks might have given rise to the low-accessibility zone in that area. The low accessibility zone to the west of Rockhampton might also be related to a local relief increase: the centroid of the less statistical area involved is situated next to Arthurs Bluff State Forest and, in the extension thereof, Blackdown Tableland National Park. These areas rise abruptly above the surrounding lowlands (Queensland Government; Department of National Parks; Recreation; Sport and Racing, 2013). As such, the cliff tops of Blackdown Tableland National Park's undulating plateau (Queensland Government; Department of National Parks; Recreation; Sport and Racing, 2013) act as a local barrier. In Australia's more inland regions, the less accessible statistical areas also partly coincide with Australia's major deserts, which are again characterised by a low(er)-density road network. Hence, an inadequate road network infrastructure may lead to extended travel times and thus lower accessibility regions, even when a well-connected or hub airport (e.g. Sydney Airport or Melbourne Airport) is situated in the vicinity of the origin location involved. At the same time, the way in which the points of origin and destinations were defined may also influence the bimodal accessibility index to a considerable degree. This is of particular interest in cases where the population weighted centroid was replaced by the geometric centroid, since this may artificially increase the distance between the centroid involved and the main road network. Hence, **Error! Reference source not found.** is not simply a map of 'major airports', but indicates the combined effects of land- and airside connectivity on the accessibility of locations.

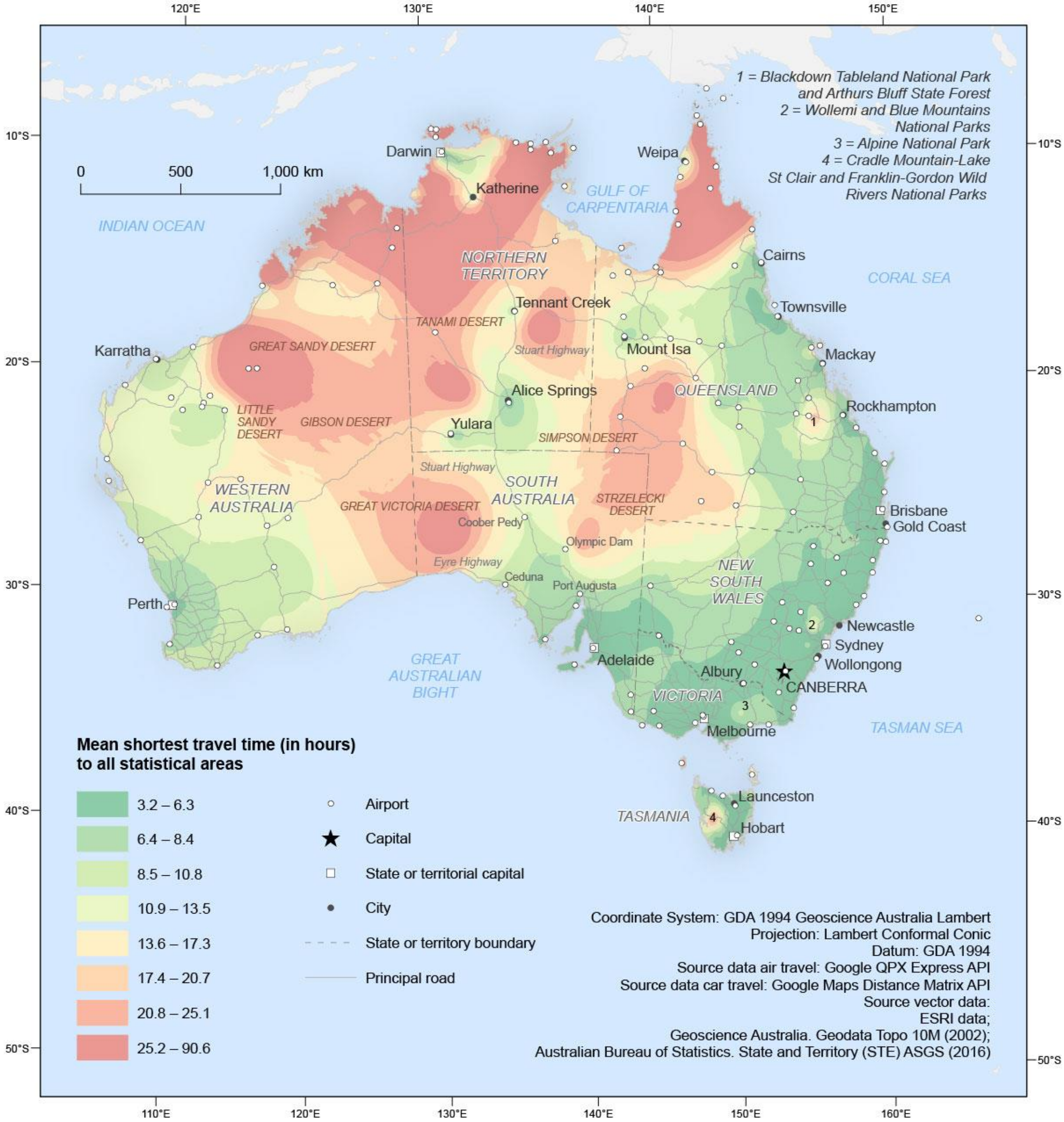

Figure 5 : The bimodal accessibility pattern within Australia. (Adapted from Meire *et al.*, in press)

5. Discussion

Our case study shows some of the possibilities of gathering web-based data accessed via web APIs for accessibility related research topics. In this respect, a number of strengths of such an automated API-based approach to assessing transport networks can be identified. A major strength of the Google Maps API involves the provision of recent and/or real time information (Hajinasab et al., 2017; Wang & Xu, 2011). According to Wang & Xu (2011), for example, the Google Maps service is generally updated twice a month, which is likely to be significantly more than, for example, static road network shapefiles for GIS network analysis. Second, although not adopted in our study, the Google Maps API provides a diversity of traffic configurations among which different traffic models, potentially producing additional insights. Third, the outsourcing of the travel data collection itself reduces the need for self-preparing a transport network (e.g. in a GIS environment), which contributes to a faster data collection process (Hajinasab et al., 2017; Wang & Xu, 2011). Finally, a major strength of such an API-based approach involves the public availability of Google's travel APIs. Although the use of Google's APIs is charged according to the amount of requests sent, no restrictions apply on who may access and use Google's API services. However, we recognise that the use of web-based travel data, accessed via web APIs, also contains a number of limitations: the dependency on corporate web-based data might raise both operational and ethical concerns. Google's QPX Express API service, for example, has been ended as from April 10, 2018. The reproducibility of our research approach might therefore be questioned. However, provided that the programming code is adjusted to an alternative travel API, our framework for assessing the combined land- and airside accessibility can be considered a generic approach, which can be applied in similar research. Ethical concerns in turn relate to the selective openness and limited transparency of corporate (web-based) data. Related to this, the reliability of web-based travel data, and in this study Google-based data in particular, should be evaluated.

6. Concluding remarks

In this study, based on Meire et al. (in press) and Meire and Derudder (under review), we introduced a framework for transport data extraction and processing using publicly available web-based resources (i.e. the Google Maps Distance Matrix and QPX Express APIs), illustrated by a case study evaluating the combined land- and airside accessibility of Australia at the level of statistical units. The results of our case study have demonstrated some of the main possibilities, strengths and weaknesses of gathering web-based data via web APIs for transport and accessibility related research topics. A main avenue for future research involves a detailed quality assessment of Google's web-based travel data, including the operational and ethical issues related to this. With regard to our case study, future research could focus on further developing the bimodal accessibility index (e.g. by incorporating complementary transport variables such as traffic conditions) and monitoring how and to which extent the accessibility pattern changes over time. In conclusion, we argue that the development of new data gathering and processing methods in the realm of big data offers new and exciting opportunities for transport research in general and accessibility research in particular.

7. References

Ahmadi, M., Dileepan, P., & Wheatley, K. K. (2016). A SWOT analysis of big data. *Journal of Education for Business*, *91*(5), 289–294. https://doi.org/10.1080/08832323.2016.1181045

Australian Airports Association. (2012). *Australia's Regional Airports: Facts, Myths, & Challenges*.

Australian Bureau of Statistics. (2011). *Australian Statistical Geography Standard (ASGS): Volume 5 - Remoteness Structure*. *Australian Bureau of Statistics* (Vol. 5). Sydney. Retrieved from http://www.abs.gov.au/AUSSTATS/abs@.nsf/DetailsPage/1270.0.55.005July 2011?OpenDocument

Australian Bureau of Statistics. (2016a). Australian Population Grid 2016 in ESRI Grid Format. Retrieved from http://www.abs.gov.au/AUSSTATS/abs@.nsf/DetailsPage/3218.02015-16?OpenDocument

Australian Bureau of Statistics. (2016b). Statistical Area Level 2 (SA2) ASGS Ed 2016 Digital Boundaries in ESRI Shapefile Format. Retrieved from https://www.abs.gov.au/AUSSTATS/abs@.nsf/DetailsPage/1270.0.55.001July 2016?OpenDocument

Barceló, J., Montero, L., Marqués, L., & Carmona, C. (2010). Travel Time Forecasting and Dynamic Origin-Destination Estimation for Freeways Based on Bluetooth Traffic Monitoring. *Transportation Research Record: Journal of the Transportation Research Board*, (2175), 19–27. https://doi.org/10.3141/2175-03

Bassett, L. (2015). *Introduction to JavaScript Object Notation*. Sebastopol, CA: O'Reilly Media, Inc.

Becker, R. A., Cáceres, R., Hanson, K., Loh, J. M., Urbanek, S., Varshavsky, A., & Volinsky, C. (2011). A Tale of One City: Using Cellular Network Data for Urban Planning. *IEEE Pervasive Computing*, *10*(4), 18–26. https://doi.org/10.1109/MPRV.2011.44

Cui, J., Liu, F., Janssens, D., An, S., Wets, G., & Cools, M. (2016). Detecting urban road network accessibility problems using taxi GPS data. *Journal of Transport Geography*, *51*, 147–157. https://doi.org/10.1016/j.jtrangeo.2015.12.007

Demissie, M. G., Correia, G., & Bento, C. (2015). Analysis of the pattern and intensity of urban activities through aggregate cellphone usage. *Transportmetrica A: Transport Science*, *11*(6), 502–524. https://doi.org/10.1080/23249935.2015.1019591

Derudder, B., & Witlox, F. (2005a). An appraisal of the use of airline data in assessing the world city network: A research note on data. *Urban Studies*, *42*(13), 2371–2388. https://doi.org/10.1080/00420980500379503

Derudder, B., & Witlox, F. (2005b). On the use of inadequate airline data in mappings of a global urban system. *Journal of Air Transport Management*, *11*(4), 231–237. https://doi.org/10.1016/j.jairtraman.2005.01.001

Donehue, P., & Baker, D. C. (2012). Remote, rural, and regional airports in Australia. *Transport Policy*, *24*, 232–239. https://doi.org/10.1016/j.tranpol.2012.08.007

Dumbliauskas, V., Grigonis, V., & Barauskas, A. (2017). Application of Google-based Data for Travel Time Analysis: Kaunas City Case Study. *PROMET - Traffic&Transportation*, *29*(6), 613–621. https://doi.org/10.7307/ptt.v27i2.1487

Fuellhart, K., Derudder, B., O'Connor, K., & Zhang, W. (2015). Geographic correlates of lowest available airfares on Australian air routes. *Journal of Air Transport Studies*, *6*(2), 1–22.

García-Albertos, P., Picornell, M., Salas-Olmedo, M. H., & Gutiérrez, J. (n.d.). Exploring the potential of mobile phone records and online route planners for dynamic accessibility analysis. *Transportation Research Part A*. https://doi.org/10.1016/j.tra.2018.02.008

González, M. C., Hidalgo, C. A., & Barabási, A.-L. (2008). Understanding individual human mobility patterns. *Nature*, *453*, 779–782. https://doi.org/10.1038/nature06958

Google Company. (2018a). Facts about Google's acquisition of ITA Software. Retrieved March 20, 2018, from https://www.google.com/press/ita/faq.html

Google Company. (2018b). Google Maps Platform: Developer Guide. Retrieved March 26, 2018, from https://developers.google.com/maps/documentation/distance-matrix/intro

Grubesic, T., Horner, M., Zook, M., & Leinbach, T. (2006). Global Distribution Systems and the U.S. Commercial Air Industry: Gathering Real-Time Airline Flight and Fare Information for Spatial and Economic Analysis. In *Transportation Research Board 85th Annual Meeting (No. 06-0547).*

Grubesic, T., & Zook, M. (2007). A ticket to ride: Evolving landscapes of air travel accessibility in the United States. *Journal of Transport Geography, 15*(6), 417–430. https://doi.org/10.1016/j.jtrangeo.2006.12.002

Gu, W., Wang, X., & McGregor, S. E. (2010). Optimization of preventive health care facility locations. *International Journal of Health Geographics*, *9*(17), 1–16. https://doi.org/10.1186/1476-072X-9-17

Hainen, A. M., Wasson, J. S., Hubbard, S. M. L., Remias, S. M., Farnsworth, G. D., & Bullock, D. M. (2011). Estimating Route Choice and Travel Time Reliability with Field Observations of Bluetooth Probe Vehicles. *Transportation Research Record: Journal of the Transportation Research Board*, (2256), 43–50. https://doi.org/10.3141/2256-06

Hajinasab, B., Davidsson, P., Holmgren, J., & Persson, J. A. (2017). On the use of on-line services in transport simulation. *Transportation Research Procedia, 21*, 208–215. https://doi.org/10.1016/j.trpro.2017.03.090

Jagadish, H. V. (2015). Big Data and Science: Myths and Reality. *Big Data Research*, *2*(2), 49–52. https://doi.org/10.1016/j.bdr.2015.01.005

Kung, K. S., Greco, K., Sobolevsky, S., & Ratti, C. (2014). Exploring Universal Patterns in Human Home-Work Commuting from Mobile Phone Data. *PLoS ONE*, *9*(6), e96180. https://doi.org/10.1371/journal.pone.0096180

Malinovskiy, Y., Saunier, N., & Wang, Y. (2012). Analysis of Pedestrian Travel with Static Bluetooth Sensors. *Transportation Research Record: Journal of the Transportation Research Board*, (2299), 137–149. https://doi.org/10.3141/2299-15

Meire, S., Derudder, B., & Ooms, K. (in press). A Bimodal Accessibility Analysis of Australia's Statistical Areas. *Journal of Maps*.

Meire, S., & Derudder, B. (under review). A Bimodal Accessibility Analysis of Australia Through Application Programming Interfaces.

Mogaji, E., & Erkan, I. (2019). Insight into consumer experience on UK train transportation services. *Travel Behaviour and Society*, *14*, 21–33. https://doi.org/10.1016/j.tbs.2018.09.004

Niu, Q., Wang, Y., Xia, Y., Wu, H., & Tang, X. (2018). Detailed assessment of the spatial distribution of urban parks according to day and travel mode based on web mapping API: A case study of main parks in Wuhan. *International Journal of Environmental Research and Public Health*, *15*(8). https://doi.org/10.3390/ijerph15081725

Nutley, S. (2003). Indicators of transport and accessibility problems in rural Australia. *Journal of Transport Geography*, *11*(1), 55–71. https://doi.org/10.1016/S0966-6923(02)00052-2

Padeiro, M. (2018). Geographical accessibility to community pharmacies by the elderly in metropolitan Lisbon. *Research in Social and Administrative Pharmacy, 14*, 653–662. https://doi.org/10.1016/j.sapharm.2017.07.014

Queensland Government; Department of National Parks; Recreation; Sport and Racing. (2013). *Blackdown Tableland National Park Management Statement 2013*. Retrieved from https://parks.des.qld.gov.au/managing/plans-strategies/statements/blackdowntablelands.%0Ahtml.

Rashidi, T. H., Abbasi, A., Maghrebi, M., Hasan, S., & Waller, T. S. (2017). Exploring the capacity of social media data for modelling travel behaviour: Opportunities and challenges. *Transportation Research Part C: Emerging Technologies*, *75*, 197–211. https://doi.org/10.1016/j.trc.2016.12.008

Redondi, R., Malighetti, P., & Paleari, S. (2011). The accessibility of European regions and

airport network. In *15th ATRS World Conference, Sydney, June 29 -July 2, 2011*.
Ryerson, M. S., & Kim, A. M. (2018). A drive for better air service: How air service imbalances across neighboring regions integrate air and highway demands. *Transportation Research Part A*, *114*, 237–255. https://doi.org/10.1016/j.tra.2017.10.005
Serna, A., & Gasparovic, S. (2018). Transport Analysis Approach Based on Big Data and Text Mining Analysis From Social Media. *Transportation Research Procedia*, *33*, 291–298. https://doi.org/10.1016/j.trpro.2018.10.105
Socharoentum, M., & Karimi, H. A. (2015). A comparative analysis of routes generated by Web Mapping APIs. *Cartography and Geographic Information Science*, *42*(1), 33–43. https://doi.org/10.1080/15230406.2014.976656
Stipancic, J., Miranda-Moreno, L., Labbe, A., & Saunier, N. (2017). Measuring and visualizing space–time congestion patterns in an urban road network using large-scale smartphone-collected GPS data. *Transportation Letters*, 1–11. https://doi.org/10.1080/19427867.2017.1374022
Sweigert, A. (2015). *Python: Automate the Boring Stuff*. San Francisco, CA: No Starch Press, Inc. Retrieved from https://automatetheboringstuff.com/
Tranos, E., & Mack, E. (2018). Big data: A new opportunity for transport geography? *Journal of Transport Geography*, *In press*. https://doi.org/10.1016/j.jtrangeo.2018.08.003
Versichele, M., de Groote, L., Claeys Bouuaert, M., Neutens, T., Moerman, I., & Van de Weghe, N. (2014). Pattern mining in tourist attraction visits through association rule learning on Bluetooth tracking data: A case study of Ghent, Belgium. *Tourism Management*, *44*, 67–81. https://doi.org/10.1016/j.tourman.2014.02.009
Wang, F., & Xu, Y. (2011). Estimating O-D travel time matrix by Google Maps API: Implementation, advantages, and implications. *Annals of GIS*, *17*(4), 199–209. https://doi.org/10.1080/19475683.2011.625977
Witlox, F. (2015). Beyond the Data Smog? *Transport Reviews*, *35*(3), 245–249. https://doi.org/10.1080/01441647.2015.1036505
Wong, R. C. P., Szeto, W. Y., Wong, S. C., & Yang, H. (2014). Modelling multi-period customer-searching behaviour of taxi drivers. *Transportmetrica B: Transport Dynamics*, *2*(1), 40–59. https://doi.org/10.1080/21680566.2013.869187
Xia, N., Cheng, L., Chen, S., Wei, X. Y., Zong, W. W., & Li, M. C. (2018). Accessibility based on Gravity-Radiation model and Google Maps API: A case study in Australia. *Journal of Transport Geography*, *72*, 178–190. https://doi.org/10.1016/j.jtrangeo.2018.09.009
Zhang, S., & Zhou, W. (2018). Recreational visits to urban parks and factors affecting park visits: Evidence from geotagged social media data. *Landscape and Urban Planning*, *180*, 27–35. https://doi.org/10.1016/j.landurbplan.2018.08.004
Zuo, T., Wei, H., & Rohne, A. (2018). Determining transit service coverage by non-motorized accessibility to transit: Case study of applying GPS data in Cincinnati metropolitan area. *Journal of Transport Geography*, *67*, 1–11. https://doi.org/10.1016/j.jtrangeo.2018.01.002