

5

Tracking via Geotagged Social Media Data

What this chapter will cover:

- How geotagged social media data offers researchers the opportunity to collect vast amounts of big data from some, but not all, social media platforms.
- The variety of methods that can be employed to obtain geotagged social media data.
- The range of conceptual insights that have emerged from this cost efficient and reliable form of data.
- That the non-covert manner of this form of research requires careful steps to ensure individuals' data is treated in an ethical manner.

Introduction

Over the past twenty years, social media has changed the ways in which we plan, travel and reflect on our travels. Tourists use social media while travelling to stay in touch with friends and family, enhance their social status (Guo et al., 2015); and assist others with decision making (Xiang and Gretzel, 2010; Yoo and Gretzel, 2010). They also use it to report back to their friends and family where they are. This can be done

using a geotag function that provides a location for where a post is made. While little is known about why tourists choose to geotag their social media posts, Chung and Lee (2016) suggest that geotags may be used in an altruistic manner by tourists, in order to provide information, and because they elicit a sense of anticipated reward. What is known, however, is that the function offers researchers the ability to understand where tourists travel.

There are two types of geotagged social media data. The first of these is discussed in this chapter and may be defined as single point geo-referenced data – geotagged social media posts whose release is chosen by the user. This includes data gathered from social media apps such as Facebook, Instagram, Twitter and WeiChat. The method of obtaining this data involves the collation of large numbers of discrete geotagged updates or photographs. Data can be collated via an application programming interface (API) provided by the app developer to researchers, by automated data scraping via computer programs, perhaps written in Python, or manually by researchers. The second type of data is continuous location-based data from applications that are designed to track movement constantly, such as Strava or MyFitnessPal. Tracking methods using this continuous location-based data are discussed in detail in the following chapter.

Geotagged social media data can provide rich information for planners and tourism managers, as well as for tourists (Garcia-Palomares, Gutierrez and Minguez, 2015). The use of data from this source is often referred to as crowd sourcing, digital foot printing, or data crawling (Mukhina, Rakitin and Vishertin, 2017). It is an extremely useful tool to indicate the level of attractiveness – or unattractiveness – of different spaces (Kachkaev and Wood, 2013). Further, it can be done in real time. However, it may also be regarded as a fast-moving method that is subject to ethical constraints and constraints related to platforms' terms and conditions. This chapter will review this method, its application within tourism, and the ethical and legal considerations that are associated with this technique.

Using crowd sourced geotagging to track tourist mobility

In the past ten years, there has been an explosion of publicly available geotagged social media data on the internet, as a result of apps offering the geotagging functionality to its users. Facebook, Instagram, Snapchat and Twitter are examples of social media apps that offer this functionality. Geotagging is defined as a process whereby geospatial information, temporal information, and/or textual information is added to online content such as written posts or photographs (Dickinger et al., 2008; Kádár and Gede, 2013; Wong, Law, Li, 2017). This is possible because individuals' devices (mobile phone, computer or tablet) add their location to text, photographs or documents – and this can be done manually by the user each time they post, or automatically via their settings within the app (Dickinger et al., 2008). The majority of research in this space has focussed on geotagged photographs or text, but research that assesses geotags can include articles and video content.

As mentioned earlier, the use of data produced through web-based platforms is often referred to as crowdsourcing. Crowdsourced data has been defined by Walden-Schreiner et al. (2018: 782) as:

'... information that is generated by many individuals, often accessed through web platforms'. Data generated from this method can include '... geotagged photos shared publicly on social-media and other social-media-derived sources of volunteered geographic information (i.e., GPS tracks).'

Since its early use by authors such as Girardin et al. (2008), the collation of geotags as a tool to explore tourism behaviour has emerged as a significant method through which tourist mobility and behaviour may be understood (Walden-Schreiner et al., 2018). Data sourced using this method can also assess the economic impact of tourism (Sonter et al., 2016) and tourists' preferences for biodiversity (Hausmann et al., 2017). Ironically, despite its enormous potential, Wong, Law and Li (2017: 48) describe the field as being in the 'nascent stage' with the majority of research being published in the information technology or computer science field.