Big Data and Cycling

DOI:10.1080/01441647.2015.1084067

Gustavo Romanillos[a*], Martin Zaltz Austwick[b], Dick Ettema[c] & Joost De Kruijf[d]

Author affiliations

[a] tGIS Transport, Infrastructure and Territory Research Group, Complutense University of Madrid, Profesor Aranguren S/N, Ciudad Universitaria, 28040 Madrid, Spain

[b] Bartlett Centre for Advanced Spatial Analysis, University College London, Gower Street WC1E 6BT, London

[c] Faculty of Geosciences, Utrecht University, PO Box 80115, 3508 TC Utrecht, The Netherlands

[d] Faculty of Geosciences, Utrecht University, Heidelberglaan 2, Room 621, 3584 CS Utrecht, The Netherlands

ABSTRACT

Big Data has begun to create significant impacts in geography, urban and transport planning. This paper covers the explosion in data-driven research on cycling, most of which has occurred in the last ten years. We review the techniques, objectives and findings of a growing number of studies we have classified into three groups according to the nature of the data they are based on: *GPS data (spatiotemporal data collected using the Global Positioning System), live point data,* and *journey data.* We discuss the movement from small-scale GPS studies to the "Big GPS" datasets held by fitness and leisure apps or specific cycling initiatives, the impact of Bike Share Programmes (BSP) on the availability of timely point data and the potential of historical journey data for trend analysis and pattern recognition. We conclude by pointing towards the possible new insights through combining these datasets with each other - and with more conventional health, sociodemographic, or transport data.

Keywords: Cycling, Big Data, bike mobility, bikeshare, spatial analysis, GPS

1. Introduction

Big Data holds the promise to illuminate social processes that were previously undersampled or poorly understood. For those involved in city planning, service provision, and business intelligence, it still remains central to innovation and research. The term arose first from the large-scale collective efforts of scientists at the CERN (*Conseil Européen pour la Recherche Nucléaire*) particle accelerator, large scale astronomy and genomics projects (Marx, 2013) – but for more than five years, the potential for working with large-scale social data has been grasped by the commercial sector (Manyika, 2011) as well as governments and non-governmental organisations (NGOs) (Hall, 2012). Despite the excitement it has generated, working definitions of the term are problematic – the most widely adopted framework derived from Laney (2001) refers to the "3Vs" of Big Data: Volume (size), Velocity (speed of generation or collection) and Variety (synthesizing a range of sources). Later authors (Kitchin, 2014) have added additional definitions to this (including "Veracity", the quality of the data – as a way to preserve the alliteration of the concept), but it seems dubious that, in the wider world of Big Data, many data sources fully qualify under all the categories of the 3Vs, or the wider definitions. Most of the data sources discussed in this review qualify as Big Data under the first V (Volume), but possibly not the others – many are single source (e.g. a transport provider or single app or web platform, disqualifying them under the *variety* criterion) and few provide large velocities of data in real-time.

It perhaps makes sense to view the concept of Big Data as representing an enthusiasm for the rapid expansion of data availability. Within these technologically-driven definitions, there is no focus on openness or accessibility., While the promise of innovation and new markets may motivate engineers and computer scientists, it is the availability of data that has empowered and excited new actors in policy, politics and governance. New datasets have become widely accessible which capture the detail of processes that previously were estimated, under sampled, kept private, or simply poorly understood. In part, the Open Data movement can be thanked for its hand in not only pushing an agenda of transparency, but encouraging service providers and government departments to provide usable datasets and streaming APIs (Application Programme Interfaces) that third parties can use to create commercializable platforms and research outputs. The topics of data released as a result of a movement towards Open Government Data (OGD) arguably has antecedents in census and administrative data, and the transparency agenda has driven the release of largely pre-existing datasets (see, for example, Coleman (2013)). However, the presence of technology as a mechanism of automation and monitoring has generated new datasets with collection methods which are distinct from centrally-compiled or volunteered OGD. This is particularly true in transport, where the automated systems for ticketing or charging create a uniquely detailed data stream – however, this data stream has significant enough privacy issues that it's not yet available in this detailed form. Transport and geolocated data has quite an incredible capacity to de-pseudononymise and reveal new information about individuals (for example, the work done on open data around New York taxis to 'stalk' celebrities or identify the homes of people who go to strip clubs (Tockar, 2014)), so there is a very clear rationale for caution about open data release in this sphere.

Perhaps the most notable example of this data boom is the expansion of smart card systems for public transport in major cities (Pelletier, Trépanier & Morency, 2011) providing journey level information for individual users, in systems that were previously sampled by gate counts and travel-to-work questionnaires. The quantum leap from limited to almost complete sampling is unprecedented, and time slices of this data are available to researchers or developers through service providers online (for example, Transport for London (2014)). Cycling sits in a nexus where availability of Big Data (from quantified self data, BSP, GPS devices and mobile tracking) intersects with societal needs around fitness, sustainability and air quality, and service provision and infrastructure planning for active transport.

This review seeks to survey the Big Data sources available to cycling researchers, broadly split into *GPS data, live point data,* and *journey data.* These data follow different patterns of volume and velocity, suggesting different problem domains and generating differing analysis approaches. GPS data is collected via smartphone, embedded devices, or specialized units – this is usually collected by individual users within the context of a quantified lifestyle (using fitness, health, and leisure apps), or contributing to a specific study. While this could be shared and acted upon in real time, in many cases users will upload their route at the end of a journey or at the end of the day, putting it in the category of historical data. These provide a high level of data density. Typical GPS data is sampled every few (three to five) seconds, generating hundreds of data points per individual journey, and depending on the sample period, thousands per user, and hundreds of thousands or millions in a typical GPS study (for example, Hood, Sall & Charlton, (2011)). In the case of fitness apps and social media-driven systems, this can number tens of millions of users and routes (Endomondo, 2013; Map My Ride, 2014). Working with GPS data poses some challenges with respect to accuracy (Schuessler & Axhausen, 2009a) and volume, but it has also been one of the more fruitful in terms of the application of models which can link directly to transport planning policy on a city or county level.

Point data refers to information collected at a particular location – an example of this is the information provided by a docking station in a BSP (Froehlich, Neumann & Oliver, 2009), or the data transmitted by a traffic camera or gate counter at a specific intersection (Rogers & Papanikolopulos, 2000). This tends to be smaller in volume, but the increasing availability of this data is starting to allow some extensive insights such as the research conducted by O'Brien, Cheshire, & Batty (2013), which analysed 38 BSP located in Europe, Asia, the Middle East, Australia and the Americas. Furthermore, through web APIs, BSP can provide information in real time for immediate analysis and response. The rich spatiotemporal characteristics of this data have led to some novel applications of cluster analyses.

Journey data acts at a coarser level than GPS data – providing origin and destination locations and times for individual journey, but not necessarily including detailed information about route choice, actual link speed and delay. A number of bikeshare programmes (BSP) have released journey data covering a period of months, often amounting to several million journeys – but at present, with some exceptions such as the

Capital Bike Share initiative (2015), this data is released months after the fact, making it more amenable to long-term trend analysis than nowcasting or rapid response. The origin-destination datasets allow for space-time and network approaches, and researchers have used route inference to generate the spatial richness of GPS tracks on multi-million journey scale (Zaltz Austwick, O'Brien, Strano, & Viana, 2013), although few estimates of the robustness of these inferences have been carried out.

2. Research focused on GPS data

Global Positioning System (GPS) technology was originally developed in the 1970s, but despite being available for civil purposes in the mid-1980s, it was only in the 1990s that it became widespread in its integration into consumer devices (Kumar & Moore, 2002). Since then, GPS data have been collected for transport analysis (Shen & Stopher, 2014). Initially, the technology was mainly applied to improve aerial and maritime navigation systems, but since the late-1990s the largest application of GPS has been land transport. Over the last twenty years GPS data has been collected for evaluating system performance such as measuring historical congestion and flow levels, analysing travel behaviour and estimating route choice models (Rasmussen, Ingvardson, Halldórsdóttir, & Nielsen, 2013). In the field of mobility, GPS data have also been collected in the context of household travel surveys, in order to complement the survey responses with detailed trip reporting for a subset of journeys (Bricka, Sen, Paleti, & Bhat, 2012; Doherty, Noel, Gosselin, SIROIS, & UENO, 2001; Ohmori, 2005; Shen & Stopher, 2014).

Since 2007 there has been a substantial rise in the volume of GPS data, due in part to the smartphone "revolution". In 2009 smartphones accounted for 15.4% of the general pool of mobile phones (Li et al., 2010), by 2014 it surpassed 35%, with over 175 billion units (eMarketer, 2014). In the US, it rose from 44% in 2011 to 65% in 2013 (The U.S. Digital Consumer Report, 2014). The generalised presence of GPS technology in smartphones and the vast growth of mobile applications based on location and tracking functionalities also fed this growth. The emergent navigation and the sport/fitness app markets (Evans, 2013; Flurry Analytics, 2014) became apparent more recently, linking personal recorded data to online platforms where people can display and manage their routes and information, and share and compete with other people, creating different user-communities.

In this section we focus on bicycle riding GPS data collected through mobile applications, GPS devices and online platforms specifically created for each study, and data from big app companies, only recently available for research and planning purposes.

2.1. GPS data collected through specific research initiatives

The first work analysing cycle mobility through GPS tracks dates from 2007 (Harvey & Krizek, 2007). In spring 2006, the research team launched an initiative to recruit volunteers from different neighbourhoods in South Minneapolis, and finally collected 938 trips from 51 participants (selected according to their age, gender, home location and work location) using GPS devices in order to study commuter cyclist behaviour, analysing chosen routes and their variations due to existing bike facilities. The project remarked on the difficulty of cleaning GPS data, which can contain significant positional inaccuracies - consequently, analysis of cycling behaviour is improved by mapping the recorded GPS tracks onto street infrastructure. Different authors (Wagner, 1997; Marchal et al., 2005 and Schuessler & Axhausen, 2009a) determined diverse approaches to the map-matching process that, with increasing complexity and sophistication, solved the main problems.

The work of Harvey and Krizek provided a descriptive approach to cyclist behaviour. Subsequent studies focused on developing cyclist route choice models from larger samples of GPS routes – typically studying thousands of cyclists and their routes. The first of these studies, conducted in Zürich (Menghini, Carrasco, Axhausen, & Schüssler, 2010), analysed nearly 2500 journeys from over 2400 cyclists. The sample size allowed the creation of a route choice model, but, since this research did not collect any data associated with the cyclists or the trips, disaggregation by individual and important features of the street network (such as slope or traffic) were omitted in the model.

The sample analysed in Zürich was obtained from an independent GPS study that collected raw GPS data from nearly 5,000 participants carrying a GPS receiver for up to a week, resulting in over 32 000 trips in the cities of Zürich, Winterthur and Genève. The raw data was processed to identify different transport modes and trips (Schuessler & Axhausen, 2009b), extracting cycle journeys for independent analysis. Modes were detected based on the average and maximum speed during the trip, or by investigating vicinity to infrastructure and stations/stops during the trip. In the latter case, geo-data regarding stops and infrastructure was linked to the GPS data using Geographic Information Systems (GIS). For instance, Stopher et al. (2008) first extract walking trips, followed by public transport trips. Of the remaining trips, bicycle trips were extracted based on speed and acceleration characteristics. They comment that GPS loggers can be configured such that they will not record when stationary (to save the battery). However, when the respondent starts moving again the logger needs some time (up to a few minutes) to locate its position, potentially leading to missing trip starts, which requires additional pre-processing. Broach, Dill & Gliebe (2011, 2012) developed a route choice model from GPS data collected in Portland, Oregon, focussing on the journeys of regularly commuting cyclists. This was a smaller study (with only 164 subjects and around 1500 trips), but its small scale allowed the research team to collect more detailed demographic data via questionnaire – recognising that cyclists are a heterogeneous community whose route choices might vary significantly.

At approximately the same time, in Los Angeles, Reddy et al. (2010) had carried out the first study using smartphones as a mechanism for collecting GPS data. With the aim of building a platform that enriched the route sharing process, the *Biketastic* project developed a mobile application for Android phone users and distributed it online for free, recruiting 450 users (Savage, 2010). The project website allowed participants not only to visualise and manage their trips and statistics, but also to share their routes, and visualise other cyclist's journeys and other data. GPS data was associated with noise level and roughness data collected through the smartphones' microphones and accelerometers. Volunteers could also provide information about the route as well as uploading photos and videos of the journeys – acting as a community resource, but also providing contextual data for researchers.

Similar schemes followed in San Francisco, California (Hood, Sall, & Charlton, 2011), and Austin, Texas (Hudson, Duthie, Yatinkumar, Larsen, & Meyer, 2012). The first of these used the mobile application *CycleTrack,* developed for the study by Charlton, Schwartz, Paul, Sall, & Hood (2010) and made available for Android and Apple iOS in an effort to broaden the volunteer base. The initiative collected the largest sample of cycle GPS tracks to date for research purposes, with nearly one thousand volunteers contributing data over a five-month period. Through the app, volunteers provided data about their gender, age and travel purpose, which were incorporated into the route choice model. Unfortunately, fewer that one third of these journeys were successfully mapped to the road network for further analysis. This cleaning and map-matching processing was improved by the research conducted shortly afterwards using the same GPS smartphone application in Austin, Texas (Hudson et al., 2012). Although a smaller study, they succeeded in matching a similar number of routes. In both of these studies, the participants were recruited from the smartphone users community, raising the question of sample bias; however, comparing demographic data from the smartphone study with information obtained from local travel surveys did not reveal significant difference in mean age, although they did reveal a gender bias towards males in the smartphone study. Other socio-demographic data, such as income, were not collected to avoid private concerns. Smartphone ownership might have a skew in that regard, but it has not been possible to test this.

Following these pioneering studies, more recent research initiatives have focussed on smartphone GPS applications, improving the online platforms and websites that link apps with volunteers, and providing new functionalities to encourage people to participate. The initiative *Madrid cycle track* (Romanillos, 2013; 2014) engaged three hundred casual bikers, as well as cyclists for bike-messenger companies. The initiative collected over 45 000 km of GPS tracks through a free mobile application, *Map My Tracks*. In an effort to broaden the user base, those without smartphones had the option of drawing their routes on an online map. In both cases, associated information about the age and gender of participants and the purpose of the travel was collected. It was also the first initiative to allow volunteers to visualise the whole network of collected tracks on a single online map.

In the Netherlands, a similar community-focussed initiative was created to generate interest in pedelecs (electric bicycles). *B-Riders* in Noord-Brabant in the Netherlands started in September 2013 and ended December 2014, with the aim of shifting users from car travel to pedelec use. Participants could either register for a financial compensation - from €0.10 to €0.15 for each kilometre registered in the morning or the evening peak hours, with a limit of €1,000 for each participant, or register for a coaching program with feedback and encouragement on their individual behaviour, or both. To receive the financial compensation and the feedback, participants were obliged to make use of a smartphone GPS application developed for the program – resulting in an unprecedented 400 000 GPS tracks collected over the period. Bike Print (2014), which allows visualisation and summary of the data by users (such as specific length of the trip), was developed specifically for the task, and the data was subsequently used to predict future usage of the bike network (Coevering, Leeuw, Kruijf, & Bussche, 2014).

2.2. Big GPS Data from "big app" companies

The volume of GPS data collected by studies increased significantly when researchers implemented GPS mobile applications. The development of associated online platforms, and advertising campaigns among the cyclist community, served to engage larger groups of participants. However, the sample of contributors still tends to be small compared to the cycling population in the studied locations. The growth in sports and fitness apps have opened up sampling of huge numbers of users (Evans 2013; Flurry Analytics, 2014). In the US nearly one-third of smartphones owners (46 million people) currently use health or fitness apps (Nielsen, 2014a), aided in part by smart watches and fitness bands (Nielsen, 2014b). These wearable devices are however currently mostly appealing and affordable for a limited group of wealthy young people, and even within this group, two thirds of users do not use these devices for more than six months (Mitesh, Patel, MBA, & Hall, 2015). Among these fitness apps, GPS sports tracking apps have been especially popular. In 2013, 7 of these apps surpassed 16 million downloads (Comstock, 2013); in 2013, the popular *Endomondo* celebrated its fifth birthday and reached 20 million users in more than 200 countries (Endomondo, 2013). *MapMyFitness* experienced an even more rapid expansion, surpassing 20 million members in October 2013 (Map my fitness, 2014). App developers ascribe this popularity to attractiveness of the social dimension of the service as well as the introduction of new features like training plans (Endomondo, 2013). We are living in the era of not only Big Data, but Big Apps.

These apps are widely used by cyclist for tracking sport activities. *Endomondo* has registered almost a billion miles of cycling activities, more than half of the total uploaded (Endomondo, 2013). *MapMyRide*, one of the most popular together with *Strava*, has over 20 million users (Map My Ride, 2014), who have uploaded over 70 million routes (My fitness pal, 2014). *Strava* does not disclose its number of users, but 2.5 million GPS-tracked activities are uploaded to its website every week (Strava, 2014a) and more than 90 million rides have been collected (Albergotti, 2014).

There are limited studies on these new big GPS datasets from app companies. Cintia, Pappalardo & Pedreschi (2013) examined GPS tracks of nearly 30 000 cyclists, collected via the *Strava* API and analysed training performance using average speed, duration of ride and cyclist's heart rate. Wamsley (2014) focussed on analysing travel times collected through Strava in order to generate pacing strategies for a cyclist to complete a course in the fastest time possible. Other research defined the conceptual architecture of data collection, management and methodologies for using and analysing the data (Clarke & Steele, 2011), including data cleaning, visualisation and trajectory clustering techniques (Peixoto and Xie, 2013). Other work has instead focussed on the use, the motivations and the online community experience for the people that use cycling apps (Smith, 2014). Very few researchers in this field have focussed on the analysis of urban transport cycling to improve urban planning and design (Clarke & Steele, 2011) or have developed specific tools to analyse cyclists' routes. Researchers in Reykjavik (Jónasson, Eiriksson, Eðvarðsson, Helgason, & Sæmundsson, 2013) have done work in this area, using GPS data from *Garmin Connect* and *Strava* online platforms to create heat map and analyse cyclist route choices .

The research and planning disciplines are traditionally more interested in urban transport cycling and require high data density, and data which is representative of the population in their study region, to build and validate models which big app data does not necessarily provide. This is beginning to change, as *Strava* is the first of these companies to sell cycling GPS data. On May 2014, the company launched *Strava Metro*, a commercial brand of the company focussed on providing data services to local authorities, research institutions, and other interested parties (Strava Metro, 2014a). In 2013 (Maus, 2014), Oregon's Department of Transportation was the first partner to sign with *Strava* (Albergotti, 2014). Other urban planning authorities around the world (including London and Glasgow in the UK, and Victoria in Australia) have followed suit (Albergotti, 2014; Sparkes, 2014). Strava have also launched *Strava Labs*, a high-resolution online map that visualises the cycle flow distribution collected through the app around the world (Strava Labs, 2014), representing over 75 million journeys and 220 billion GPS points (Mach, 2014).

Models like *Strava Metro* bring significant new opportunities for analysis and understanding. First, the *Street map* shows a very high density of GPS tracks covering the whole metropolitan area (although still exhibiting some degree of spatial and sociodemographic bias). The data is processed to remove users' personal information, but summaries of basic demographic information (gender and age ranges) *are* provided, allowing demographic bias to be estimated. Additionally, it provides not only information about the total number of cycle trips but also the number of commuting trips - very important information for urban transport planning. *Strava Metro* also provides cyclist flow information at different dates and times – e.g. via the *Strava Saturday* online heat map (Strava, 2014b)- so it is possible to analyse cyclist flow for different times of the day (the morning and the afternoon peaks), and study the evolution across the whole year, opening up the possibility of detailed spatiotemporal and seasonal analyses.

However, *Strava Metro* data also presents limitations. Users' privacy concerns mean that single route tracks are typically not accessible so it's not possible to analyse trip length, purpose of travel or the route choice on an individual journey level. Because this data is shared in an aggregated form, it is not possible to study the relationships between these variables; for example, the dependence of route choice on the cyclist's travel purpose. Because we only have aggregated socio-demographic information, there is limited scope to analyse the importance of basic factors like age or gender in route planning, journey length or purpose. All of these analyses are likely to be important for planning, designing and managing cycle infrastructure. The solution would be to have access to disaggregate data and provide single tracks, a difficult proposition when maintaining user (and company) privacy. In order to not discourage user participation, shortly after opening *Strava Metro*, the company offered members the option of marking routes as private – these routes are then not included in *Strava Metro* dataset (Wehner, 2014).

3. Research focused on point data

As well as the substantial body of research around GPS, there has been a significant interest in analysing cycling data gathered at specific locations. Studies have mainly explored two different data sources: point data registered at Bike Share Programmes (BSP) stations and counts.

3.1. Exploring Bike Share Programmes data mines

With the exception of studies based on bike parking data provided by specific, one-off surveys (Rietveld, 2000), bike mobility trends have not been analysed through large point datasets gathered at BSP docking stations or parking lots - until recently. The biggest evolution in this area came with the rapid expansion of BSP in cities around the world. The *first generation* of such systems date from 1965 (Demaio, 2009), but they remained very few and small in size till the early-1990s, when a *second generation* of BSP was born. Still these programs grew slowly until the mid-2000s, when a *third generation* of bike share (characterised by electronic management, and hence a rich data source) became popular in many countries. Since then, the number of such systems increased exponentially around the world (Fishman, Washington, & Haworth, 2013). By the end of 2007 there were about 60 cities with third generation BSP implemented worldwide (Demaio, 2007); according to Fishman (2015) the current number of BSP is 855, with nearly one million bicycles in use.

A common feature of this third generation of BSP is that they record information when a bike in undocked (hired) or docked (returned). This data was first explored in a study in the Barcelona BSP, *Bicing* (Froehlich et al., 2009), covering August to December 2008. Three different kinds of data were gathered from the *Bicing* information system by scraping the website (using an automated program to find and store the relevant data elements presented by the webpage). This data was collected every two minutes and included the station locations, the number of available bicycles, and the number of vacant parking slots. *Bicing* launched in 2007; it had nearly 400 stations and 6,000 bikes, with 150 000 subscribers. Firstly, by applying clustering techniques, the research identified spatiotemporal patterns, relating the use of different bike stations to activity clusters over the course of a weekday, when more regular BSP usage patterns were identified. Secondly, the research developed different predictive models to analyse the impact of several factors (such as time of the day or the amount of historical data) in order to create tools to estimate bicycle demand for different stations and the optimal location of future ones. The research pointed towards the potential of this new source of data to identify not only cycling or mobility patterns, but broader urban trends and dynamics, such as inferring urban land uses (home, office or leisure/retail) by analysing users' profile over time.

A later study worked with Barcelona BSP data with more specific objectives (Kaltenbrunner, Meza, Grivolla, Codina, & Banchs, 2010). Aware that users of *Bicing* often found it difficult to find a bike to hire, or a space to leave their bike at their destination, the researchers developed a model that could predict the availability of bikes or docks, and could inform both users and system managers in advance so that they could respond accordingly. Even an hour ahead, their autoregressive–moving-average (ARMA) model was typically accurate to one bicycle, representing a usable prediction range for cyclists. More recently, Giot & Cherrier (2014) completed a similar predictive analysis based on Washington, D.C. BSP data, working with a suite of research regression techniques.

There has been a range of effort to work with BSP data in real time, building new tools for system management and to improve service. In 2009 Luo & Shen (2009) developed an information system for the BSP of Hangzhou (China) that represented the location of the BSP stations and dynamically displayed the availability of bikes or free parking spots. The most remarkable visualisation of real time BSP information is *The Bike Share Map*

(O'Brien, 2010; 2013). Created in 2010 in order to visualise London's BSP data, the map represents the information of different cities around the globe since June 2013, covering at time of writing 107 BSP and visualising the availability of systems around the world. This global view was incorporated into research based on BSP data (Cheshire & O'Brien, 2013; O'Brien, Cheshire, & Batty, 2013). The investigation collected data from 38 systems from Europe, the Middle East, Asia, Australia and America, and the dataset included locations, capacity and current load factor of docking stations. After analysing the data, the investigation compared and classified the BSP according to variables such as the system's geographical size, the variation of occupancy rates across the day or the week, and the intensity and distribution of activity in relation to demographics. The paper compared the geographical distribution and temporal popularity of a range of different schemes, allowing planners to examine schemes with elements in common in other parts of the world.

As well as research focussing on providing useful apps and interfaces to service providers, researchers are increasingly taking more theoretical approaches to dock data to understand differing spatiotemporal patterns using signal processing and statistical methods. In 2012, Lathia, Ahmed and Capra, (2012) used cluster analysis to detect "similar" stations in the London system based on the time profile of their occupation, resulting in docking stations which have similar behaviours over the course of a day, and examining the impact of "casual" users. These users pay using a credit card instead of the access keys used by subscription users at the time of the programme's launch - these casual users may be more likely to be tourists or business visitors . Similar methods were applied by Côme & Latifa (2012) to cluster docking stations which are similar in their temporal patterns of occupation, focussing on the flagship *Velib'* system in Paris. This covered 2.5 million trips in just one month - *Velib'* is the second largest BSP in the world. Working on the London system, Padgham (2012) is one of the first to attempt to connect BSP activity with that of the other parts of the public transport network, and introduced spatial interaction model-like approaches to understanding flows between locations. Many of these studies focussed on Europe and North America. Corcoran, Rohde, Charles-Edwards & Mateo-Babiano (2014) studies Brisbane, Australia and examines the impacts of weather and public events on city cycle use. In Melbourne, Fishman, Washington, Haworth and Mazzei (2015) used data collected from BSP trips in 2012 to visually represent the strength of the relationship between different docking stations and how this relates to the public transport system

Research on point data in BSP systems has yielded a raft of visualisations, apps and analyses. Many of the more academic works have employed specialised statistical techniques that are perhaps not as familiar to the policymaker or transport planner, and joining up the scientific expertise with services and interventions amenable to the user, service provider or policymaker still has a way to go. Limited work has been done to combine it with journey data, which in itself would yield new possibilities.

3.2. Other point data sources: Manual and automated counts

While BSP provides detailed and timely point data reporting, there are other sources that provide large and useful point data collections, but rarely on the same scale and level of detail. Within the scope of this review, the evolution of counts in the last years is especially interesting.

Though manual counts cannot be considered as a source of Big Data – they just meet the first V criterion (volume) of Laney's (2001) classification - they are still the most prevalent cycling data collection method (Ryus, Laustsen, Proulx, Schneider, & Hull, 2014), producing increasingly large datasets through recent initiatives. Many communities still successfully use conventional, lower-tech methods in order to collect point data and support an evidence base for cycling policy. In some countries, like the US, many cycling communities (Schneider, Patten, & Toole, 2005) encourage volunteers to register cyclists at key locations in precise dates through manual count methods. Among the different initiatives, especially remarkable is the *National Bicycle and Pedestrian Documentation Project* (NBPD, 2009-2015), a program that provides to the volunteers a methodology, as well as training and documentation, and centralises the collection of surveys and counts from cities all around the US.

Apart form these massive manual counts initiatives, there is a substantial collection of cycling data through automated counts. The most common methods are based on pneumatic tubes, inductive loops, passive infrared, automated video counters, infrared cameras and fiber optic pressure sensors (Ryus et al., 2014). Pneumatic and inductive are widespread, but proved to be accurate only when detectors are properly installed, calibrated, maintained, free of external interference, and on a dedicated bicycle lane (Nordback & Janson, 2010). Recently, more innovative counts based on fiber optics register cyclists on mixed traffic lanes, offering insight not only in the cycling volume but also in the speed and direction. In the Netherlands, new traffic light detection loops have been implemented to detect cyclists with high accuracy by using a new methodology with dedicated algorithms (Winter, 2012; Rijn, 2014). This system is being implemented extensively in some cities: Utrecht is currently adjusting 170 traffic lights which measure motorised traffic to also detect cyclists. This cycling data is being made available in an online open data platform (Open Data Utrecht, 2015). Such efforts could be facilitated by the technological innovators who are working to create sensors which cost close to $50 – 1% of the cost of current sensors (Andersen, 2015). *Knock Software* is one such innovator, active in Portland, OR on a device which uses magnetic, thermal and speed detection to determine whether a passing object is a bike, a car or a pedestrian. If this proves reliable, coverage of cities could rapidly become more comprehensive, detailed and timely.

Considering that count data is at the base of many studies which examine travel patterns, it is worthy to highlight the most important advantages and disadvantages in relation to other approaches. Count data register every single cyclist at a specific location while BSP or GPS data relies on a more segregated cycling population. However, the absence of sample bias in count data is not guaranteed at all, and it is collected on an aggregate level such that no demographic data is captured. According to Ryus et al. (2014), manual counting is still the most dominant method of counting cyclists - 87% of total counts in the US - and still relies heavily on volunteers. That means that samples are usually registered at a limited number of locations in a specific date or period of time, and may have spatial bias if the count locations are not well distributed. The increasing extension of new automated counts could allow pattern analysis across time - and, if well distributed, could reduce spatial biases.

4. Research focused on journey data from Bike Share Programmes

The third generation of BSP not only record information about the number of bicycles in docking stations, but also identify and register bikes (and sometimes an identifier for their users) at the start and end dock of every journey. This means that BSP are able to provide general mobility data through the origin-destination matrices associated with users', but also timings of these journeys (and, by inference, duration). In addition, BSP may provide data about cyclists (age or gender, for instance) – although this is not always the case, either because the data is not collected (from casual, credit card users), or because that aspect of the data is withheld for privacy reasons. Research on journey data has so far been more limited. BSP journey data is historical; it is typically released in large batches covering months or even years of activity. It has limited use for nowcasting or feeding back information to users in real time. Nevertheless, there has been significant work in visualising this data (Wood, 2011; Zaltz Austwick et al., 2013; Bargar, Gupta, Gupta, & Ma, 2014), creating a comparison study of different visualisation techniques with respect to this data.

The research carried out by Borgnat et al. (2011) is one of the first analytical approaches to these origin-destination datasets, and focussing on data from the city of Lyon in France. The investigation analysed the dataset provided by the managing company and the City Hall, corresponding to the 13 million trips over a two and half year period. The system registered the start time and departure station, and end time and destination station, for each journey. For the first time, researchers could examine individual mobility, characterising different groups according to the distance, duration or speed of their trip. While the research carried out in Barcelona on point data (Froehlich et al., 2009), covered a short period of time, the research conducted in Lyon allowed trend and temporal analysis over a much longer period. The data collection began at the opening of the system and covered expansions of the scheme, allowing the study to cover different demand and service scenarios throughout this period, and analysed how factors such as increasing numbers of bicycles and stations affected the number of subscribers. The same year, Vogel, Greiser, & Mattfeld (2011) analysed similar data from Vienna's BSP, *Citibike Wien*, covering around 760 000 rides from 2008 and 2009. General spatio-temporal patterns are derived from the analysis while an integrated approach of Data Mining and Operation Research is presented in order to develop a new trip model that anticipates bike activities for better long-term location planning. The researchers were able to formulate clear policy goals from their analyses.

The first multi-city analysis of origin-destination data was carried out by Zaltz Austwick et al. (2013), which compared five cities (London, Washington DC, Minneapolis, Denver and Boston), using spatial network analysis methods to cluster stations into communities (subnetworks of journeys within the wider network). The smallest of these datasets covered 168 000 journeys (Denver) and the largest 3.6 million (London) and allowed comparison of distance travelled and journey time distributions between cities. The paper also used inferred routing for visualisation purposes using Open Street Map and Routino (http://routino.org), but did not utilise this for distance estimation or street network loading, as there was no mechanism to validate this route choice. Bargar et al. (2014) builds on a network analysis approach (examining data from Washington DC, Chicago and

Boston), complementing it with the spatiotemporal clustering methods used by other researchers, and visualising both of these techniques via a web-based map visualisation built using JavaScript libraries, integrating analysis into a more accessible visualisation tool.

More recent work has expanded its scope beyond predicting demand or detecting similar locations, and has focussed instead on correlating cycling activities with wider policy goals around health and transport. The use of the London BSP across the three first years of operation have been examined by Goodman & Cheshire (2014). The study analysed the evolution in the profile of users, the increase in the number of trips as well as variation in the proportion of trips by registered users. This covered a period of time that included the extension of the BSP network in 2012 and the rise of the service prices in January 2013. The dataset incorporated the gender and home postcodes of users, permitting analyses that linked geographic socio-economic factors of the residential locations, and evaluating the demand according to the distance from homes to the start or end stations. Defined as "trips made by two or more cyclists together in space and time" data (Beecham & Wood, 2014, p.1), group-cycling journeys on London BSP were studied by analysing the trips of over 80 000 members between September 2011 and September 2012. The research revealed some plausible patterns, like the increase of group cycling journeys at weekends, late evenings and lunchtimes, and the large proportion of group members that share the same postal code. However, it also revealed some unexpected ones, like sets of commuting group cycling journeys, and some differences between group and individual trips according to gender. This simple approach starts to connect BSP work with wider interests around social behaviour, health and leisure. Faghih-Imani, Eluru, El-Geneidy, Rabbat, & Haq (2014) studied how land use, urban form, building environment attributes and weather impact on the bicycle flow, by analysing the data from the Montreal BSP, *BIXI*, between April and August 2012. The research reports, unsurprisingly, good weather leading to high cycling flow, but also provide interesting findings for policy makers and urban designers, such as the relationship between BSP usage and urban density, and the interaction between cycling and public transport.

An underused aspect of journey data is its capability to act as a supplementary and validating data source for the more current, accessible point data (which through APIs, is typically updated on a minute-by-minute basis). Point data typically registers only net changes – so, for example, three bikes arriving and two bikes leaving appears the same way as one bike leaving. By using journey data to validate the behaviour of the system, it could be used to infer expected traffic at docking stations (and hence whether a small net change represents large or small flows), as well as allowing spatial models for predicting flows based on just the total ins and outs of each docking station (in GIS, interpolating a matrix from its marginal sums is a relatively standard technique (Deming, 1940)).

Future work on BSP will surely rely on combining different strands of data from within the scheme, or with external datasets. If BSP utilise GPS tracking more widely, it could open up the possibility of a linking of journey data (time-varying origin-destination matrices), point data (station locations and statuses) and routing data (the details of the route that users take between origin and destination on the street network) – allowing inference of time-dependent BSP traffic on the level of individual road segments. If GPS data yields route preference, and journey data yield time-dependent demand at an origin-destination level, combining both with live point data could yield a complex, timely modelling tool. This BSP "nowcasting" could allow prediction in very small time windows – for example, docking station-level occupation and demand in ten or twenty minutes in the future. Combining BSP data with complementary sources – health and demographic data, for example – opens up the possibility to linking BSP to a wider context – including transport planning, access to services of marginalised groups, and behaviour change.

5. Conclusions

This paper reviews the recent bike mobility research based on the analysis of Big Data collected from sources that are becoming increasingly accessible to researchers and policy makers, offering a panoramic view on the growing number of studies that, in less than ten years, have evolved as quickly as the data itself. Even if the achievements are remarkable, there are still important limitations that are difficult to overcome using current data sources. By some estimates, cycling data meets the first of Laney's (2001) "4Vs" classification of Big Data (that of volume), given the size GPS and BSP data, and perhaps the second criterion (Velocity), since some data is available in real time (Luo & Shen, 2009; O'Brien, 2010, 2013). It is more questionable whether the other V criteria (Variety and Veracity) are met, at least in the way that the data is currently being used. In the context of cycling, while the data is combined with demographic or interview data, pooling it with Big Data from other sources seldom occurs. As hinted, there may be scope within BSP to combine point data (sparse, complete and real-time data) with journey data (more detailed, complete and historical samples) and GPS data (very detailed but potentially smaller samples, and historical) to leverage the detail of one dataset against the timeliness and sampling power of the others.

With respect to Veracity, our conclusions differ between sources; this criterion refers to possible biases, noise or any abnormality in data, which is variable for each of the data types. Research based on dedicated GPS data collections have typically paid attention to proper sampling procedures, so that the collected data is by and large representative for the population studied. However, data from big app companies rely on volunteers uploading their cycling tracks, leading to self-selective samples. For instance, logging bike trips in Strava may be more likely to be carried out by cycling enthusiasts who wish to show off their cycling achievements. This would imply a lack of representativeness of the population in terms of cycling attitude, geographical location and socio-demographic characteristics. Groups with mobility impairements, those who are "afraid to cycle", elderly cyclists, or children may not be well-represented in these accounts. Recent studies by Buck (2013a) and Dill and McNeil (2013) demonstrate that heterogeneity along these lines indeed exists, suggesting that data from big app sources will be biased. However, BSP point and journey data is representative, at least of users of BSP. How representative this population is of wider cyclists and citizens is, of course, open to question (see Buck et al., (2013b) for further discussion). Indeed, there is no reason to believe that either BSP or big app data provides representative samples of a cities' population of cyclists or potential cyclists.

Another reason to be concerned about data veracity relates to data collection motivation and methods. In cases in which data is collected specifically for academic purposes, it is typically enriched with contextual information (such as socio-demographics, attitudes, spatial context or environment). When data is collected by commercial applications, aimed at providing a service to customers (e.g. Strava, MapMyRide), privacy policies of companies make using this contextual information difficult or impossible. As a consequence, key variables to understanding travel behaviour, such as socio-demographics or purpose of the journey, may be absent. However, the size of the data gathered and its continuity over time potentially allows for analyses not possible on dedicated GPS data (e.g. spatial clustering or the variation of cyclist flow distribution over time), which may deliver useful additional insights. Similarly, BSP data is collected for management rather than research, and lacks socio-demographic context. In any case, BSP may offer a rich database for analysing regularities in patterns of supply and demand as well as longer term structural developments.

On a technical level, GPS accuracy is not an issue which has been completely resolved. Dedicated GPS devices perform better than smartphones GPS (Lindsey, Gorjestani, Hankey, & Wang, 2013) but their lack of accuracy in some urban areas can mean analysts lack the fine detail to precisely distinguish route choice – one of the main reasons the data is of interest. The Galileo European Program, which is expected to be in place by 2019 (European Commission, 2014), promises improvements over the current system, but these improvements have yet to be fully demonstrated. For users, one barrier is that, historically, GPS apps have rapidly drained their smartphone batteries – this is significant enough that the *B-Riders* scheme developed an app for an intelligent start and end of the GPS tracking to minimise this problem.

Despite these caveats, there are interesting research challenges and opportunities from the increasing availability of new datasets and the steady improvements in their quality. The industries around sport-tracking apps have seen increases in the number of users of GPS devices (including recent wearable devices) (Nielsen, 2014a). If this trend continues, the volume of data will increase with the userbase, and, through licensing schemes, so will the availability of data. Data from BSP will likely grow, due to the proliferation of BSP around the world. Future research will have to face the challenge of bias in its data collections, and create robust, scalable mechanisms to account for it. We expect more GPS data to become available in a more timely fashion, not only from app companies (some of which are already offering this service for users, like Map My Tracks) but from the current third generation of BSP. Some recent systems record GPS tracks for every journey, which may allow researchers to analyse bike routes and improve the existing route choice and cycling flow distribution models, as well as analyse the real use of existing bike infrastructure. Apart from these improvements regarding raw location data, work is needed on enriching these data with meaningful explanatory variables. Socio-demographic data may be approximated by linking location data to usage patterns of specific groups. More work will also be needed on data fusion techniques in order accommodate such approximations; however, data providing spatial context (such as land use) is becoming increasingly accurate and more freely available. This growth in bicycle data and its corresponding availability, and joining up with data on transport, health, air quality, demographics, route choice and leisure promises a rich period of activity for researchers in all of these areas.

Finally, a key question remains: how will the expected advances benefit cyclists and potential cyclists, policy makers and BSP? And how would those benefits create wider impacts? Will they encourage more people to cycle, or reduce congestion or pollution? Many BSP users currently take advantage of real time information about the availability of bicycles in different docking stations so that they can plan their journeys. In a near future, we might imagine a smart bike route planning system, integrated in a multimodal transport system. Users will have information about the closest available station to their destination point, and about the best route possible for getting there, incorporating weather, traffic, and user preference – lowering barriers to cycling for less confident or experienced cyclists. Cycling Apps will continue to be attractive to users of smartphones and perhaps a new generation of wearable technology, providing information to cyclists and reports of their peers' performance, motivating people to cycle longer, faster, and of course, more frequent.

For policy makers, the range of benefits may be more diverse. GPS based cycling data will provide insights about cyclists' route choice behaviour and their preferred and disliked route characteristics, which will support the design of cycling infrastructure networks. Coupling GPS based cycling data with geo-data (land use, facilities, altitudes, etc.) will greatly enhance their understanding of cyclists' route choice. Big Data will drive the assessment of cycling infrastructure at different levels, analysing the use of local infrastructures (such as lanes or bike parking), identifying the main cycling routes over the course of a day, or understanding the obstacles, delays and dangers that slow or hinder their journeys. Again, a key issue here is the representativeness of the pool of GPS users. While an initiative such as the Dutch *BikePRINT* project delivers useful insights in cycling routes and cycling densities, it relies on voluntary participants, leaving questions about reliability of the outcomes (Coevering et al., 2014).

The recent collaboration between commercial Apps and planning institutions is promising and will generate combined and useful information that will make new explorations possible. As we have remarked, these new Big Data will not substitute but complement other more conventional sources, since they often lack disaggregate data on the cyclists, which are so often necessary to understanding the contexts that influence many of their decisions. This points, then, to a future where the fourth V – Variety – creates new innovations and insights in cycling – as Big App data, real-time BSP feeds, and more traditional, detailed, demographic studies are brought together – and commercial, municipal, service provision and academic partners work together to create a breathing, user-centred picture of the cyclable city.

6. References

Albergotti, R. (2014). Strava, Popular With Cyclists and Runners, Wants to Sell Its Data to Urban Planners. Wall Street Journal. Retrieved from http://blogs.wsj.com/digits/2014/05/07/strava-popular-with-cyclists-and-runners-wants-to-sell-its-data-to-urban-planners/

Andersen, M. (2015). This $50 device could change bike planning forever. Bike Portland ORG. Published on January 13th, 2015, retrieved from http://bikeportland.org/2015/01/13/50-device-change-bike-planning-forever-130891

Bargar, A., Gupta, A., Gupta, S., & Ma, D. (2014). Interactive Visual Analytics for Multi-City Bikeshare Data Analysis. In The 3rd International Workshop on Urban Computing. New York City, USA. Retrieved from http://www2.cs.uic.edu/~urbcomp2013/urbcomp2014/papers/Bargar_Bikesharing.pdf

Beecham, R., & Wood, J. (2014). Characterising group-cycling journeys using interactive graphics. Transportation Research Part C: Emerging Technologies, 1–13. doi:10.1016/j.trc.2014.03.007

Bike Print (2014). Retrieved from http://www.bikeprint.nl/index.php?lang=en

Borgnat, P., Robardet, C., Rouquier, J.-B., Abry, P., Fleury, E., & Flandrin, P. (2011). Shared bicycles in a city: A signal processing and data analysis perspective. Advances in Complex Systems, 14(3), 1–24. Retrieved from http://www.worldscientific.com/doi/abs/10.1142/S0219525911002950

Bricka, S., Sen, S., Paleti, R., & Bhat, C. (2012). An analysis of the factors influencing differences in survey-reported and GPS-recorded trips. Transportation Research Part C, 21(1), 67–88. doi:10.1016/j.trc.2011.09.005

Broach, J., Dill, J., & Gliebe, J. (2011). Bicycle Route Choice Model Developed from Revealed-Preference GPS Data. In Transportation Research Board 90th Annual Meeting.

Broach, J., Dill, J., & Gliebe, J. (2012). Where do cyclists ride? A route choice model developed with revealed preference GPS data. Transportation Research Part A: Policy and Practice, 46(10), 1730–1740. doi:10.1016/j.tra.2012.07.005

Buck, D. (2013a). Encouraging Equitable Access to Public Bikesharing Systems. ITE Journal, 83(3), pp.24–27. Retrieved from: http://faculty.washington.edu/abassok/bikeurb/resources/media/abstracts/papers/153_Buck.pdf

Buck, D., Buehler, R., Happ, P., Rawls, B., Chung, P., & Borecki, N. (2013b). Are Bikeshare Users Different from Regular Cyclists?. Transportation Research Record: Journal of the Transportation Research Board, 2387(1), 112-119.

Capital Bike Share (2015). Retrieved from http://www.capitalbikeshare.com/

Charlton, B., Schwartz, M., Paul, M., Sall, E., & Hood, J. (2010). CycleTracks: a bicycle route choice data collection application for GPS-enabled smartphones. In 3rd Conference on Innovations in Travel Modeling, a Transportation Research Board Conference.

Cheshire, J., & O'Brien, O. (2013). Revealing and Informing Transport Behaviour from Bicycle Sharing Systems. The Geographic Information Systems Research UK (GISRUK). University of Liverpool. Retrieved from http://www.geos.ed.ac.uk/~gisteac/proceedingsonline/GISRUK2013/gisruk2013_submission_31.pdf

Cintia, P., Pappalardo, L., & Pedreschi, D. (2013). "Engine Matters": A First Large Scale Data Driven Study on Cyclists' Performance. In 2013 IEEE 13th International Conference on Data Mining Workshops (pp. 147–153). Ieee. doi:10.1109/ICDMW.2013.41

Clarke, A., & Steele, R. (2011). How Personal Fitness Data Can be Re-used by Smart Cities. In Seventh International Conference on Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP) (pp. 395–400).

Coevering, P. van de, Leeuw, G. de, Kruijf, J. de, & Bussche, D. (2014). Bike Print. Policy Renewal and Innovation by means of Trakcing technology. In Nationaal verkeerskundecongres 2014.

Coleman, E. (2013) Lessons from the London Datastore. Goldstein, B. & Dyson, L. (Ed.) Beyond Transparency: Open Data and the Future of Civic Innovation (pp. 39-50).

Côme, E., & Latifa, O. (2012). Model-based count series clustering for Bike-sharing system usage mining , a case study with the Vélib ' system of Paris . Transportation Research-Part C, 1–23.

Comstock, J. (2013). 7 fitness apps with 16 million or more downloads. Mobi Health News. Retrieved from http://mobihealthnews.com/24958/7-fitness-apps-with-16-million-or-more-downloads/

Corcoran, J., Li, T., Rohde, D., Charles-Edwards, E., & Mateo-Babiano, D. (2014). Spatio-temporal patterns of a Public Bicycle Sharing Program: the effect of weather and calendar events. Journal of Transport Geography, 41, 292-305.

Demaio, P. (2007, December 30). What a year for the bike-sharing. Retrieved from http://bike-sharing. blogspot.com/2007/12/what-year-for-bike-sharing.html/.

Demaio, P. (2009). Bike-sharing : History , Impacts , Models of Provision , and Future. Journal of Public Transportation, 12(4).

Deming, W.E. & Stephan, F.F. (1940). On a least squares adjustment of a sampled frequency table when the expected marginal totals are known. The Annals of Mathematical Statistics, 11(4), 427–444.

Dill, J. & McNeil, N. (2013). For types of cyclists? Examining a typology to better understand bycycling behavior and potential. Proceedings of the 92nd Annual Meeting of the Transportation Research Board, p.18.

Doherty, S. T., Noel, N., Gosselin, M.-L., Sirois, C., & Ueno, M. (2001). Moving beyond observed outcomes: Integrating Global Positioning Systems and interactive computer-based trave behavior surveys. Transportation Research Board.

eMarketer. (2014). Smartphone Users Worldwide Will Total 1.75 Billion in 2014. eMarketer. Retrieved from http://www.emarketer.com/Article/Smartphone-Users-Worldwide-Will-Total-175-Billion-2014/1010536

Endomondo. (2013). Endomondo Fitness App Runs Past 20 Million Users and Reaches Profitability. Retrieved from http://blog.endomondo.com/2013/10/16/endomondo-fitness-app-runs-past-20-million-users-and-reaches-profitability/

European Commission. (2014). No Title. Retrieved from http://ec.europa.eu/enterprise/policies/satnav/galileo/index_en.htm

Evans, B. (2013). Mobile is eating the world. Retrieved from http://ben-evans.com/benedictevans/2013/11/5/mobile-is-eating-the-world-autumn-2013-edition

Faghih-Imani, A., Eluru, N., El-Geneidy, A. M., Rabbat, M., & Haq, U. (2014). How land-use and urban form impact bicycle flows: evidence from the bicycle-sharing system (BIXI) in Montreal. Journal of Transport Geography. doi:10.1016/j.jtrangeo.2014.01.013

Fishman, E., Washington, S., & Haworth, N. (2013). Bike Share: A Synthesis of the Literature. Transport Reviews, 33(2), 148–165. doi:10.1080/01441647.2013.775612

Fishman, E. (2015). Bikeshare: A Review of Recent Literature. Transport Reviews, 1–22. doi:10.1080/01441647.2015.1033036

Flurry Analytics. (2014). Health and Fitness Apps Finally Take Off, Fueled by Fitness Fanatics. Retrieved from http://www.flurry.com/blog/flurry-insights/health-and-fitness-apps-finally-take-fueled-fitness-fanatics#.VA2Zmfl_tik

Froehlich, J., Neumann, J., & Oliver, N. (2009). Sensing and Predicting the Pulse of the City through Shared Bicycling. In Twenty-First International Joint Conference on Artificial Intelligence (IJCAI-09) (pp. 1420–1426).

Giot, R., & Cherrier, R. (2014, December). Predicting bikeshare system usage up to one day ahead. In Computational Intelligence in Vehicles and Transportation Systems (CIVTS), 2014 IEEE Symposium on (pp. 22-29). IEEE. Retrieved from http://ieeexplore.ieee.org/xpl/login.jsp?tp=&arnumber=7009473&url=http%3A%2F%2Fieeexplore.ieee.org%2Fxpls%2Fabs_all.jsp%3Farnumber%3D7009473

Goodman, A., & Cheshire, J. (2014). Inequalities in the London bicycle sharing system revisited: impacts of extending the scheme to poorer areas but then doubling prices. Journal of Transport Geography, 41, 272-279.

Hall, W., Shadbolt, N., Tiropanis, T., O'Hara, K., & Davies, T. (2012). Open data and charities. Retrieved from: http://eprints.soton.ac.uk/341346/

Harvey, F. J., & Krizek, K. J. (2007). Commuter Bicyclist Behavior and Facility Disruption. Transportation Research Board. Retrieved from http://trid.trb.org/view.aspx?id=811576

Hood, J., Sall, E., & Charlton, B. (2011). A GPS-based bicycle route choice model for San Francisco, California. Transportation Letters The International Journal of Transportation Research, 3(1), 63–75. doi:10.3328/TL.2011.03.01.63-75

Hudson, J. G., Duthie, J. C., Rathod, Y. K., Larsen, K. A., & Meyer, J. L. (2012). Using smartphones to collect bicycle travel data in Texas (No. UTCM 11-35-69).

Jónasson, Á., Eiríksson, H., Eðvarðsson, I., Helgason, K. T., Sæmundsson, T., Sigurgeirsson, D. B., & Vilhjálmsson, H. H. Optimizing expenditure on cycling roads using cyclists' GPS data. School of Computer Science, Reykjavik University. Retrieved from: http://trauzti.com/files/urban-routing.pdf

Kaltenbrunner, A., Meza, R., Grivolla, J., Codina, J., & Banchs, R. (2010). Urban cycles and mobility patterns: Exploring and predicting trends in a bicycle-based public transport system. Pervasive and Mobile Computing, 6(4), 455–466. doi:10.1016/j.pmcj.2010.07.002

Kitchin, R. (2014). Big data should complement small data, not replace them. Impact of Social Sciences. The Philosophy of Data Science (series). Retrieved from: http://blogs.lse.ac.uk/impactofsocialsciences/2014/06/27/series-philosophy-of-data-science-rob-kitchin/

Kumar, S., & Moore, K. B. (2002). The evolution of global positioning system (GPS) technology. Journal of science Education and Technology, 11(1), 59-80.

Laney, D. (2001). 3D data management: Controlling data volume, velocity and variety. META Group Research Note, 6.

Lathia, N.; Ahmed, S. & Capra, L. (2012) "Measuring the Impact of Opening the London Shared Bicycle Scheme to Casual Users." Transportation Research Part C: Emerging Technologies 22 (88–102. doi:10.1016/j.trc.2011.12.004.

Li, X., Ortiz, P. J., Browne, J., Franklin, D., Oliver, J. Y., Geyer, R., … Chong, F. T. (2010). Smartphone Evolution and Reuse: Establishing a More Sustainable Model. 2010 39th International Conference on Parallel Processing Workshops, 476–484. doi:10.1109/ICPPW.2010.70

Lindsey, G., Gorjestani, A., Hankey, S., & Wang, X. (2013). Feasibility of Using GPS to Track Bicycle Lane Positioning. Report CTS 13-16. Intelligent Transportation Systems Institute Center fo Transport Studies University of Minnesota. Retrieved from: http://conservancy.umn.edu/bitstream/handle/11299/148996/CTS13-16.pdf?sequence=1&isAllowed=y

Luo, R., & Shen, Y. (2009). The Design and Implementation of Public Bike Information System Based on Google Maps. 2009 International Conference on Environmental Science and Information Application Technology, 156–159. doi:10.1109/ESIAT.2009.298

Mach, P. (2014). What do 220,000,000,000 GPS data points look like? Retrieved from http://engineering.strava.com/global-heatmap/

Manyika, J. (2011). Big data: The next frontier for innovation, competition, and productivity, McKinsey & Company. Retrieved from: http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation

Map my fitness. (2014). Map my fitness. About us. Retrieved September 13, 2014, from http://about.mapmyfitness.com/about/company-history/

Map My Ride. (2014). Join over 20 million people getting fit on MapMyRide. Retrieved from http://www.mapmyride.com/routes/

Marchal, F., J. K. Hackney and K.W. Axhausen (2005) Efficient map matching of large Global Positioning System data sets: Tests on speed-monitoring experiment in Zürich, Transporta- tion Research Record, 1935, pp. 93–100.

Marx, V. (2013). Biology: The big challenges of Big Data. Nature, 498(7453), 255–260.

Maus, J. (2014). ODOT embarks on "big data" project with purchase of Strava dataset. Retrieved from http://bikeportland.org/2014/05/01/odot-embarks-on-big-data-project-with-purchase-of-strava-dataset-105375

Menghini, G., Carrasco, N., Axhausen, K. W., & Schüssler, N. (2010). Route choice of cyclists in Zurich. Transportation Research Part A: Policy and Practice, 44(9), 754–765. doi:10.1016/j.tra.2010.07.008

Mitesh, S., Patel, M. D., MBA, M. S., & Hall, B. (2015). Wearable Devices as Facilitators, Not Drivers, of Health Behavior Change. The Journal of the American Medical Association, 13(5). Retrieved from http://jama.jamanetwork.com/article.aspx?articleID=2089651

My fitness pal. (2014). Map My Ride. Retrieved September 14, 2014, from https://www.myfitnesspal.com/apps/show/184

National Bicycle and Pedestrian Documentation Project (NBPD). (2009). Fact Sheet and Status report. Retrieved Mars 15, 2015, from https://www.bikepeddocumentation.org/

National Bicycle and Pedestrian Documentation Project (NBPD). (2015). Retrieved Mars 15, 2015, from https://www. bikepeddocumentation.org/

Nielsen. (2014a). Hacking health: How consumers use smartphones and wearable tech to track their health. Nielsen. Retrieved from http://www.nielsen.com/us/en/insights/news/2014/hacking-health-how-consumers-use-smartphones-and-wearable-tech-to-track-their-health.html

Nielsen. (2014b). Tech-styles: Are consumers really interested in wearing tech on their sleeves?. Retrieved from http://www.nielsen.com/us/en/insights/news/2014/tech-styles-are-consumers-really-interested-in-wearing-tech-on-their-sleeves.html

Nordback, K., & Janson, B. (2010). Automated Bicycle Counts. Transportation Research Record: Journal of the Transportation Research Board, 2190(-1), 11–18. doi:10.3141/2190-02

O'Brien, O. (2010). The Bike Share Map. Retrieved from www.bikes.oobrien.com

O'Brien, O. (2013). Bike Share Map. Retrieved from http://oobrien.com/bikesharemap/

O'Brien, O., Cheshire, J., & Batty, M. (2013). Mining bicycle sharing data for generating insights into sustainable transport systems. Journal of Transport Geography, 34, 262–273. doi:10.1016/j.jtrangeo.2013.06.007

Ohmori, N. (2005). GPS Mobile phone-based activity diary survey. In Eastern Asia Society for Transportation Studies (Vol. 5, pp. 1104–1115).

Open Data Utrecht (2015). Retrieved Mars 15, 2015, from https://opendata.utrecht.nl/

Padgham, M. (2012) "Human Movement Is Both Diffusive and Directed." PLoS ONE 7, no. 5 e37754. doi:10.1371/journal.pone.0037754.

Peixoto, D. A., & Xie, L. (2013). Mining Trajectory Data. Retrieved from http://courses.cecs.anu.edu.au/courses/CSPROJECTS/13S2/Reports/Alves_Peixoto_Douglas_Report.pdf

Pelletier, M.-P., Trépanier, M. & Morency, C. (2011). Smart card data use in public transit: A literature review. Transportation Research Part C: Emerging Technologies, 19(4), 557–568.

Rasmussen, T. K., Ingvardson, J. B., Halldórsdóttir, K., & Nielsen, O. A. (2013). Using wearable GPS devices in travel surveys : A case study in the Greater Copenhagen Area. In Proceedings from th eAnnual Transport Conference at Aalborg University (pp. 1–26). Retrieved from http://www.trafikdage.dk/papers_2013/188_ThomasKjaerRasmussen.pdf

Reddy, S., Shilton, K., Denisov, G., Cenizal, C., Estrin, D., & Srivastava, M. (2010, April). Biketastic: sensing and mapping for better biking. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (pp. 1817-1820). ACM.

Rietveld, P. (2000). The accessibility of railway stations : the role of the bicycle in The Netherlands. Transportation Research Part D, 5, 2–6.

Rijn, B.W. van (2014) Van detectielussen naar fietsintensiteiten: Rijdend, afrijdend en roodrijders. Elst, IT&T

Rogers, S., & Papanikolopulos, N. P. (2000). Bicycle Counter. Minnesota Department of Transportation Office of Research & Strategic Services.

Romanillos, G. (2013). Huella ciclista de Madrid (Madrid Cycle Track). Retrieved from www.huellaciclistademadrid.es

Romanillos, G. (2014). Analysing and mapping the cyclable city. A GPS-based analysis of the real and potential bicycle use in Madrid. In 15th International Conference on Information Technology in Landscape Architecture.

Ryus, P., Laustsen, K. M., Proulx, F. R., Schneider, R. J., & Hull, T. (2014). Methods and Technologies for Pedestrian and Bicycle Volume Data Collection (Vol. D). Retrieved from http://onlinepubs.trb.org/onlinepubs/nchrp/nchrp_w205.pdf

Savage, N. (2010). Cycling through data. Communications of the ACM, 53(9), 16. doi:10.1145/1810891.1810898

Schneider, R., Patten, R., & Toole, J. (2005). Case Study Analysis of Pedestrian and Bicycle Data Collection in U.S. Communities. Transportation Research Record: Journal of the Transportation Research Board, 1939(-1), 77–90. doi:10.3141/1939-10

Schuessler, N., & Axhausen, K. W. (2009b). Processing Raw Data from Global Positioning Systems Without Additional Information. Transportation Research Record: Journal of the Transportation Research Board, 2105(-1), 28–36. doi:10.3141/2105-04

Shen, L., & Stopher, P. R. (2014). Review of GPS Travel Survey and GPS Data-Processing Methods. Transport Reviews, 34(3), 316–334. doi:10.1080/01441647.2014.903530

Smith, W. (2014). Mobile interactive fitness technologies and the recreational experience of bicycling: A phenomenological exploration of the Strava.

Sparkes, M. (2014). GPS Big data: making cities safer for cyclists. The Telegraph. Retrieved from http://www.telegraph.co.uk/technology/news/10818956/GPS-big-data-making-cities-safer-for-cyclists.html

Stopher, P., Clifford, E., Zhang, J., & FitzGerald, C. (2008). Deducing mode and purpose from GPS data. Institute of Transport and Logistics Studies. Retrieved from: http://ws.econ.usyd.edu.au/itls/wp-archive/itls-wp-08-06.pdf

Strava. (2014a). Does Strava have enough data to provide a meaningul dataset? Retrieved from http://metro.strava.com/thank-you/

Strava. (2014b). Strava Saturday heat map. Retrieved September 15, 2014, from http://www.strava.com/saturday-heatmap#0|12|3|30.50000|-40.80000

Strava Labs. (2014). Strava Labs. Retrieved September 15, 2014, from http://labs.strava.com/heatmap/#5/-110.69370/35.21986/blue/bike

Strava Metro. (2014). Strava Metro. Retrieved September 15, 2014, from www.metro.strava.com

The U.S. Digital Consumer Report. (2014). Retrieved September 15, 2014, from http://www.nielsen.com/us/en/insights/reports/2014/the-us-digital-consumer-report.html

Tockar, A. (2014). Riding with the Stars: Passenger Privacy in the NYC Taxicab Dataset. Retrieved from http://research.neustar.biz/2014/09/15/riding-with-the-stars-passenger-privacy-in-the-nyc-taxicab-dataset/.

Transport For London (2014). Open Data Users. Retrieved from https://www.tfl.gov.uk/info-for/open-data-users/

Vogel, P., Greiser, T., & Mattfeld, D. C. (2011). Understanding Bike-Sharing Systems using Data Mining: Exploring Activity Patterns. Procedia - Social and Behavioral Sciences, 20, 514–523. doi:10.1016/j.sbspro.2011.08.058

Wagner, D. P. (1997) Lexington area travel data collection test: GPS for personal travel surveys. Final Report, Office of Highway Policy Information and Office of Technology Applications, Federal Highway Administration, Battelle Transport Division, Columbus, September 1997.

Wamsley, K. (2014). Optimal Power-based cycling pacing strategies for Strava Segments. Kutztown University of Pennsylvania.

Wehner, M. (2014). Strava begins selling your data points, and no, you can't opt-out. Tuaw. Retrieved from http://www.tuaw.com/2014/05/23/strava-begins-selling-your-data-points-in-the-hopes-of-creating/

Winter, M. (2012) Monitoren van fietsintensiteiten in Enschede. Enschede: gemeente Enschede en IT&T

Wood, J., Slingsby, A., & Dykes, J. (2011). Visualizing the dynamics of london's bicycle-hire scheme. Cartographica: The International Journal for Geographic Information and Geovisualization, 46(4), 239-251.

Zaltz Austwick, M., O'Brien, O., Strano, E., & Viana, M. (2013). The structure of spatial networks and communities in bicycle sharing systems. PloS One, 8(9), e74685. doi:10.1371/journal.pone.0074685