

Article

Dynamic, Interactive and Visual Analysis of Population Distribution and Mobility Dynamics in an Urban Environment Using the Mobility Explorer Framework

Jan Peters-Anders ^{1,*}, Zaheer Khan ², Wolfgang Loibl ¹, Helmut Augustin ³ and Arno Breinbauer ¹

¹ Austrian Institute of Technology, Center for Energy, Sustainable Buildings and Cities Unit, 1210 Vienna, Austria; Wolfgang.Loibl@ait.ac.at (W.L.); arno.breinbauer@edu.fh-kaernten.ac.at (A.B.)

² University of the West of England, Department of Computer Science and Creative Technologies, BS16 1QY Bristol, UK; Zaheer2.Khan@uwe.ac.uk

³ City of Vienna, Municipal Department for Urban Development and Planning (MA 18), 1082 Vienna, Austria; helmut.augustin@wien.gv.at

* Correspondence: Jan.Peters-Anders@ait.ac.at; Tel.: +43-664-815-7995

Academic Editor: Willy Susilo

Received: 28 February 2017; Accepted: 25 April 2017; Published: 15 May 2017

Abstract: This paper investigates the extent to which a mobile data source can be utilised to generate new information intelligence for decision-making in smart city planning processes. In this regard, the Mobility Explorer framework is introduced and applied to the City of Vienna (Austria) by using anonymised mobile phone data from a mobile phone service provider. This framework identifies five necessary elements that are needed to develop complex planning applications. As part of the investigation and experiments a new dynamic software tool, called Mobility Explorer, has been designed and developed based on the requirements of the planning department of the City of Vienna. As a result, the Mobility Explorer enables city stakeholders to interactively visualise the dynamic diurnal population distribution, mobility patterns and various other complex outputs for planning needs. Based on the experiences during the development phase, this paper discusses mobile data issues, presents the visual interface, performs various user-defined analyses, demonstrates the application's usefulness and critically reflects on the evaluation results of the citizens' motion exploration that reveal the great potential of mobile phone data in smart city planning but also depict its limitations. These experiences and lessons learned from the Mobility Explorer application development provide useful insights for other cities and planners who want to make informed decisions using mobile phone data in their city planning processes through dynamic visualisation of Call Data Record (CDR) data.

Keywords: smart city; cell phone data records; population distribution; mobility dynamics

1. Introduction

1.1. Context and Contributions

ICT is becoming an integral part of smart city solutions to deal with a range of societal challenges including sustainable city planning. On the one hand, innovative ICT-enabled solutions provide a unique opportunity to collect and process new data to gain additional information intelligence that helps in informed decision making [1]. On the other hand, to evaluate the real potential of these new datasets requires engagement with domain experts to identify real city planning needs, develop

tailor-made solutions, conduct comprehensive experiments and perform expert validation in actual environments [2]. In the above context, this paper investigates the potential of mobile phone data (i.e., Call Data Records (CDR) that are independent of Global Positioning System (GPS) technology) to visually explore *diurnal population distribution* and *sojourn mobility patterns*. This visual exploration includes interactive spatial and temporal dynamics of urban population on a city and city-region scale. This paper documents and discusses the effectiveness of CDR in gaining new information intelligence to support urban and transport planning (e.g., rehabilitation of public spaces, reusing brown fields for new businesses, improving public transport, etc.) in selected city environments. As compared to other related work (next section), the uniqueness of this work is based on the following strengths: (i) it was driven by the requirements of real end users, i.e., urban and transport planners from the City of Vienna, which resulted in validating and evaluating the overall effectiveness of this approach in city planning; (ii) it introduces a new framework to apply CDR and land-use data in city planning; (iii) it creatively uses various visual techniques to represent temporal and spatial data in an online web system that enables end users to perform interactive and dynamic visualisation of population distribution and mobility patterns; and; (iv) it provides insights into CDR processing and highlights its strengths and weaknesses in city planning.

1.2. Background

Due to the increase in the urban population, city planners are always looking for evidence-based information sources to effectively and efficiently plan and utilise public resources. With the emergence of new ICT solutions, city planners are keen to know population dynamics in their cities, e.g., where do people spend their time (i.e., offices, homes, parks, etc.) during different hours of the day? What is the origin and destination of people travelling in a city? What is the purpose of travel, e.g., work or social? What is the mode of travel (e.g., car, bus, walking, cycling, etc.)? Can the social biases (e.g., age, gender, etc.) in population distribution be detected? How many people travel from one part (or district) of the city to another part on weekdays or weekends? What is the functional urban region of a sprawling city? Answers to these questions help planners to better plan the built environment and public transport to improve the quality of life of citizens.

Population distribution has long been detected through the census. These data are important but provide insufficient information on the real-time dynamics of population distribution in a city. This is because distribution patterns change frequently during the day and time of the year, mainly due to the daily socioeconomic activities of citizens that require them to travel from one place to another. As census data have given only a coarse image of the population distribution and its changes during the day until now, mobile phone location data (i.e., CDR) are a unique and appreciated source to collect more evidence-based information about population distribution and mobility patterns.

Every day a large amount of mobile device data is generated in almost every country around the world. As all mobile communication service providers require customer activity monitoring for technical, accounting and recently for security reasons, they log the time, location and further activity information for each mobile device, aka CDR. A new log entry is written when a device state changes because of personal actions (phoning, texting, etc.) or technical reasons (attaching to a new cell, updated cell information). These data can be used to detect patterns of population distribution and perform mobility analyses. However, the methods to do so are not widely adopted and only a small community of scientists have made use of these possibilities in recent years. The three most important reasons for this situation seem to be: (i) from public, funding agencies' and social scientists' perspectives, it is seen as untrustworthy to access these data and could also violate privacy; (ii) only a few mobile communication companies are willing (or allowed to by law) to deliver these data [3]; and (iii) the quality of the data may not be suitable to generate reliable information for planning decisions.

Mobility pattern detection refers to geographic coordinates and time stamps extracted from the mobile communication network, which is typically organised into mobile cells equipped with cell towers (bearing antennas). The technical infrastructure continuously observes the connection between

each mobile device and a respective network cell and this information is then logged. The size of these log files can be in 10s of gigabytes per day (for medium- to mega-sized cities) and requires Big Data processing approaches for smart city applications [4].

All the logged data can be useful to process and integrate with planning data to generate useful information. Such information is helpful for urban planning, urban design and infrastructure layout. Distinct time and position information of the mobile device activities as well as the customer movements allow for a mapping of the spatio-temporal distribution of the cell phone subscribers and applying them as proxy for time-specific population distribution [5]. Visualising and analysing the daily, weekly or monthly patterns of those population and mobility dynamics provide a holistic and evidence-based information source at the city scale, which can be used to support urban planning, transportation infrastructure improvement, public transport planning, open space utilisation, etc.

1.3. Problem Statement

Acquisition and processing of mobile phone data is not straightforward due to security and privacy concerns, location accuracy requirements and the large amount of data (often in gigabytes) that log the connections with their particular cell towers [6]. Also physical characteristics, e.g., city size, terrain, topography, buildings and mobile communication system properties, may vary from one city to another, which affects the process of determining mobility patterns. All these constraints require a thorough investigation, involving different stakeholders to identify city requirements and test the feasibility of mobile phone data-based population distribution and mobility analysis.

In our previous short work [3] highlighted the issues related to the acquisition of mobile data and its potential benefits in city planning. This work was complemented in [5] with basic experiments as proof of concept by using a small sample of mobile data to assess its feasibility for visualising population distribution and mobility dynamics. In this paper we go beyond feasibility tests and investigate CDR rigorously for the City of Vienna by integrating it with auxiliary city data (see chapter 4). The rigorous analyses presented here provide new insights and lessons learned about mobile phone dataset heterogeneity, generating dynamic output from a small geographical unit to covering city-region scale, data limitations and accuracy issues, pre-processing needs, geographical size of cities or terrain characteristics (e.g., mountains, landscape) that affect city planners' ability to fulfil new requirements. These new experiments were performed by developing additional features in the software tool 'Mobility Explorer' that helped in investigating the extent to which mobile phone data from a major mobile service provider is effective in smart city planning. The results are validated by domain experts and also by cross-validating against other available city data sources.

Designing and developing the Mobility Explorer was partly funded by the European Commission's FP7 Project UrbanAPI (September 2011–December 2014), aiming at supporting urban planning through ICT tools. In UrbanAPI, motion exploration using mobile phone data has been applied to three EU cities of different structure and size: (i) Vienna Region (Austria); (ii) Vitoria-Gasteiz (Spain); and (iii) Bologna (Italy). Here we will only cover the Vienna case in detail for the following reasons: (i) to fully understand the potential benefits and limitations of such data to support city planning; (ii) the obtained raw data of Vienna allow for investigating the range of possibilities of the data and its potential for city planning and the results can be replicated for other cities; and (iii) to make better use of limited space in this paper on analysis and reflection on the usage of mobile phone data for city planning. Nevertheless, data accuracy and cell antenna density issues are presented based on the Vitoria-Gasteiz case in Section 8 to reflect on the suitability of Mobility Explorer for smaller-sized cities.

1.4. Research Method

This research asserts the following hypothesis:

“Mobile phone data can be used in deriving and visualising diurnal population distribution dynamics and sojourn mobility patterns to derive necessary information intelligence for smart city planning processes”.

In order to investigate the above hypothesis, this research adopts a problem-solving approach mainly due to involvement of real case studies and end users who defined clear requirements. This approach is very close to the Design Science Research Methodology (DSRM) proposed by Hevner et al. [7,8], which is a well-accepted method in information systems research (e.g., [9,10]). Due to continuous engagement with the domain experts from cities, problem and requirements were identified and clear research motivations and objectives of the proposed solutions were defined. As a result, the Mobility Explorer framework was introduced, designed, developed, demonstrated and validated iteratively with the objective to improve its application design, CDR data structure and pre-processing algorithms. The interim evaluation also provided further inputs to improve the Mobility exploration tool and assess the feasibility of CDR for planning needs. For the development, demonstration and validation we acquired necessary mobile phone data from a major mobile phone service provider and investigated what “information” could be generated from these data. Also, additional city data (see chapter 4) were obtained from the city administration and integrated with mobile phone data to generate the required outputs. Different experiments were performed to check the accuracy and quality of mobile phone data. The outputs of the application were validated by domain experts and also by user evaluation exercises with external experts from pan-European organisations.

1.5. Paper Structure

The related work is presented in Section 2, where we highlight unique aspects of our work by comparing it to already existing publications. Then we introduce the Mobility Explorer framework in Section 3, followed by a brief introduction to Smart City data collection methods used for estimating population distribution and mobility dynamics in Section 4. Section 5 presents the Vienna case study and stakeholder requirements. Then the Mobility Explorer application’s system architecture is presented in Section 6. In Section 7 we depict the Mobility Explorer application’s data processing optimisation and visualisation. The application’s evaluation and lessons learned are presented in Section 8. Finally the paper is concluded in Section 9.

2. Related Work

The mobile phone market penetration can be observed in the European member states between 80% and 150%, e.g., in Austria 8 million citizens hold around 12 million mobile device subscriptions, resulting in a subscriber/population ratio of 150% (Source: <http://data.worldbank.org/indicator/IT.CEL.SETS.P2/countries/1W?display=default> Last accessed: 4 April 2016). Michalopoulou et al. [11] have proven that there is a spatial relationship between mobile device activities and population distribution. This suggests that the mobile device volume can be taken as proxy data, in order to spatially describe population distribution, mobile activity or motion patterns.

Analyses of mobile phone data have been conducted since the late 1990s. In the literature, there are some examples that provide good insight into the use of mobile data to derive population distribution and mobility patterns that can be used in various urban applications [12–17]. Steenbruggen et al.’s work [18] work can be considered a first step towards providing a detailed overview of the state of the art of mobile GSM data for the estimation of traffic parameters. In their work mobile phone location methods are discussed and first attempts are described in America and Europe. The findings of the Steenbruggen et al.’s article [18] provide basic research questions that can actually be answered and validated by the studies carried out in our research, as presented in this paper. For instance, the Vienna case study stretches mobile phone data (i.e., GSM data) analysis to a city and city-region scale, which provides a comprehensive estimate of mobility patterns and population distribution. This new source of information cannot be collected with conventional data collection methods, as discussed in Section 4.

Among others, a good overview of the results of mobile phone datasets is presented in Blondel et al. [19], where the authors question the suitability of the mobile data, e.g., they argue that datasets are noisy and some links appear there by chance, while others have not been captured. It would thus be interesting to question the stability of the obtained results, provided that the real network is different from what has been observed in the data. The analysis of CDR in our research verifies these concerns, and a lot of data quality checks are needed during the pre-processing of data to be able to use them within Mobility Explorer.

Similarly, Nanni et al. [20] present a case study (Ivory Coast) wherein GSM data is used to predict a transportation infrastructure demand model. The results indicate the usefulness of GSM data, which helps in deriving valid information on the systematic mobility behaviour of people between frequently visited locations for areas that lack information on mobility. Raslan and Elragal [21] present another case study on Greater Cairo where they demonstrated identification of home and work locations to predict population distribution using GSM data. In contrast, Mobility Explorer uses actual data from city databases (e.g., population census, office/work locations) with CDR to make it highly representative and to analyse the dynamics of population distributions and mobility patterns using a visual interactive tool and deduce possible reasoning.

Dash et al. [22] worked on new algorithms to predict locations (i.e., home and workplaces during weekdays and weekends) 25% more accurately using GSM data when phones are inactive. Experiments carried out in Dash et al. [23] and cross-validation of results from GSM mobile data and open statistical data provide promising results to extract population distribution patterns through an interactive application. In both the above references, the methods of predicting home and office locations are based on inactivity of mobile phones (i.e., inactive more than five hours) and regular travel patterns during weekdays, respectively, which may not generate accurate results. In Mobility Explorer CDR did not have information about inactive cell phones and hence could not be tested.

In addition, we carried out a detailed structured comparative analysis and identified commonalities and unique features developed by Mobility Explorer. This comparative analysis is shown in the matrix of Table A1 and in the comparison chart of Table A2 in Appendix A, where Mobility Explorer has been compared to other corresponding projects and publications.

3. Mobility Explorer Framework

In this section we briefly introduce a conceptual Mobility Explorer framework with the objective to highlight necessary elements that should be considered when developing a smart Mobility Explorer application for cities. Figure 1 depicts five main elements of the proposed framework.

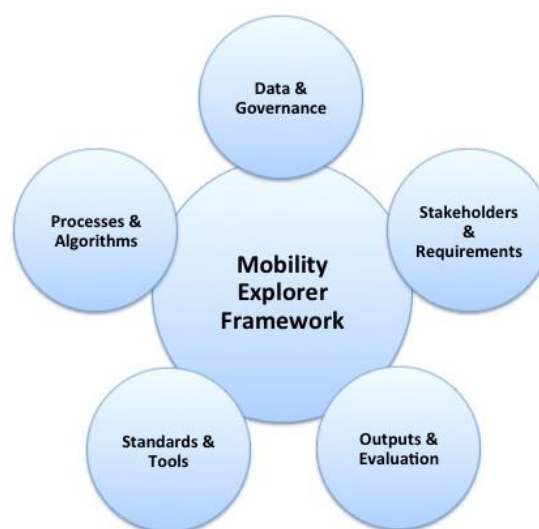


Figure 1. The Mobility Explorer framework elements.

The first element, *Stakeholders & Requirements*, emphasises the need for user-defined solutions. Stakeholder engagement and requirements development identifies planning needs and real problems in cities. These stakeholders can be city planners (e.g., urban and transport planners), data providers (e.g., citizens), mobile phone vendors, open data providers, city departments, decision makers (e.g., politicians), local businesses (e.g., real-estate developers) and others (e.g., researchers, environmentalists, health practitioners, etc.). The analyses of these users' need help to identify the necessary datasets required to develop appropriate solutions. The second element, *Data & Governance*, covers what datasets are available through different departments of cities. What auxiliary data are available from third parties? In what format is that data available? What security and privacy measures need to be taken to avoid any accidental violation of data privacy and security? How are these data accessible, etc.? These datasets can be land-use (e.g., road networks, buildings, etc.), CDR (e.g., GSM data) public surveys (e.g., local survey data), population census (e.g., demography), cross-disciplinary data (e.g., air quality, health, green spaces, etc.), social network data and other sources (e.g., GPS, traffic registers, open data, etc.). In order to assess whether the available dataset is sufficient to fulfil the city needs, the third element, *Processes & Algorithms*, defines certain alignment processes to map available datasets onto user requirements and procedures to access, secure and use datasets for processing. This further requires selecting suitable data processing algorithms (e.g., statistical analyses, machine learning, spatial and temporal analyses, etc.) to generate the expected results. The fourth element, *Standards & Tools*, necessitates adopting the system development and quality standards required by city administrations. This includes creating interoperability between new applications and legacy systems or tools by developing the necessary system interfaces so that city stakeholders can reuse application outcomes. The fifth element, *Outputs and Evaluation*, defines expected formats of the application outputs that may be in raw data form, e.g., csv files or tabular data, or in visual form, e.g., graphs, static maps or interactive maps, etc. These outputs not only help to cross-validate these results but also facilitate application evaluation to assess its benefits and usefulness for planning processes of a city. In the following sections we will cover the above elements of the proposed framework with suitable examples.

In order to operationalise the Mobility Explorer framework, in the next section we will cover data and governance aspects followed by stakeholder and requirements through the Vienna case study. Then, the standards and tools element is briefly covered through the Mobility Explorer system architecture. After that we cover *Outputs and Evaluation* together with *Processes* in detail to highlight the strengths and weaknesses of the Mobility Explorer application and its associated data.

4. Population and Mobility Data Collection Methods—Data and Governance Element

For smart cities, new data sources provide great potential to gain insight about a city's socioeconomic dynamics, which—with conventional methods—is either difficult or very expensive to get at the neighbourhood, city or city-region scale. These new insights provide evidence-based information for better planning and decision making. There are various conventional methods that have been used by cities to collect population distribution and mobility data. Most of them are expensive and provide only estimated results. These methods are indicated in Table 1.

As mentioned in Section 1.2, the mobile device volume can be taken as proxy data to detect dynamics of population distribution and big mobile communication service providers possessing large market share of subscribers may be highly representative. For Austria, CDR data collected for Vienna from the largest mobile communication service provider supplied around 5.5 million subscribers in 2015, resulting in a market share of around 42% (Source: https://www.rtr.at/de/inf/TKMonitor_1_2015/TM1_2015.pdf (page 16) Last accessed: 4 April 2016). The abovementioned numbers prove the satisfactory representativeness of the dataset to describe the entire population distribution dynamics, assuming that the mobile device distribution and motion dynamics patterns of large mobile communication companies matches quite well with the population distribution and motion dynamics patterns (see also Section 7).

Wesolowski et al. [24], Frias-Martinez et al. [25] and Blondel et al. (2015) [19] have observed that differential mobile ownership biases do not seem to have much effect on mobility patterns within particular regions. This suggests that the social bias of the clients of certain providers is a diminishing issue mainly because of highly competitive mobile services market and attractive packages. Hence little difference in subscriber behaviour patterns can be expected, at least between those of the large communication providers that ensure sufficient representativeness for mobility analysis based on mobile phones.

Table 1. Conventional data collection methods for population distribution and mobility patterns.

Data Collection Method	Benefits	Limitations
Population Census and Register Data: This is the most common method applied across the world. Population censuses or surveys are used to collect this data.	Census data are highly representative. It is not a sample but a survey covering the entire population.	In different countries the collected content varies. Also, migration and commuting to work destinations are collected through static surveys. However, census surveys are now often replaced by register counts.
Counting and Sensor Data: This static data collection method uses traffic counters or special sensors to measure traffic volume on specific roads.	It is the best estimated information about traffic volume on a specific point on the traffic network.	It does not provide any information about origins or destinations. With additional effort and under certain surrounding conditions, number plate tracking or ticket-based tracking on underground lines provides at least vague information about origins and destinations.
Polling: This is another type of survey that is used to collect mobility data.	It gives good information about origins and destinations indicated by participants. Also, it can provide additional background information about purpose of travel and reasons why a certain means of transport is chosen.	It is never representative at high spatial resolution due to limited sample sizes.

From a technical perspective, the location detection of a cell phone is done using different methods, e.g., triangulation and cell tower aggregation. Since the movement of customers causes the motion of their mobile devices, the mobile devices frequently send/receive signals to stay connected with the cell towers. If the signal quality declines, the network redirects the mobile phone automatically to a neighbouring cell that provides better signal quality. The volume of mobile devices connected to a cell tower is restricted to a certain number because of data volume limitations. Thus, in areas where a larger user number is expected, cell towers are built more densely, resulting in smaller cell extents. The cell sizes can vary from a diameter of a few 100 m in city centres to several kilometres in rural areas (Example of cell tower maps in Austria is accessible from www.senderkataster.at Last accessed: 4 April 2016). As in urban areas the cells are smaller, the location accuracy is sufficient for detailed spatial activity pattern analysis using these mobile phone location data.

However, there may be incorrect location logs due to certain reasons. For instance, cell load-balancing may result in incorrect location updates, and signal oscillation between cell towers may indicate pseudo-movements of users that actually do not happen. Such erroneous records might cause problems for detailed investigations, which have to be considered in the mobility analysis. Erroneous data issues are discussed in Section 7.

The location and time stamp information is stored continuously only for active mobile devices (of users talking, texting, e-mailing and web-browsing). The locations of mobiles in “standby” mode are just observed by the technical infrastructure, e.g., through infrequent request by the network to a device (similar to a “ping” in computer networks) to investigate if it is still turned on and where it is located. This location update between cells displays final user movement. Today the logged coordinates usually report the location of the cell tower when passing the cell border but not the exact position of the user, which is a (minor) disadvantage but can be neglected if the cells are small, which is mostly the case in dense urban environments.

Table 2 summarises mobile phone data collection methods with the benefits and limitations. These data sources can provide complementary mobility information to city planners. These methods are indicated in Table 2.

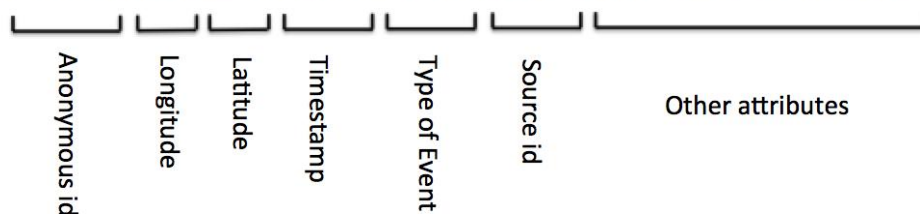
Table 2. Mobile phone data collection methods for population distribution and mobility.

Data Collection Method	Benefits	Limitations
Cell or Cell Triangulation Data: This method uses triangulation of cell towers to collect mobile phone location data.	This data collection method works passively. In fact, it is an analysis of existing log files of mobile phone service providers. This ensures a huge number of observations and thus a good representativeness even at high spatial resolutions. Also, it provides a very high temporal resolution.	It has limited spatial accuracy depending on cell size and distribution. Unfortunately, this does mean that origins and destinations of travel can only be identified roughly. Also, travel behaviour is difficult to identify because of the limited spatial accuracy; the speed of travel cannot be derived properly. This is why mode detection or detection of stops (i.e., traveller's trip chains) does not work in cities.
Global Positioning System (GPS) Data: This method uses GPS devices to collect location data.	It provides high spatial and temporal resolution.	This data collection method works actively. This means each user you want to get data from is required to install a tracking app. Supposedly it is very difficult to acquire a sufficient number of users due to privacy concerns, the battery consumption of the app and a lack of added value for the individual user. As a result, the representativeness of the data, particularly at high spatial resolutions, is very poor.
Cell Data Records (CDR): This method is used to collect data based on cell phone usage or activities.	This data collection method works passively and end users do not need to install any new app that results in battery consumption. Privacy concerns can be handled by cell service providers by anonymisation techniques. Also, the data are highly representative at high spatial and temporal resolution.	Approximation of population distribution and mobility patterns, cell tower density may vary from city to city, social biases may not be representative, non-active cell phones and more than one cell phone may generate erroneous data.

The amount of data collected by mobile service providers is enormous and for large telecom operators it can be on the order of a few hundreds of GBytes to TBytes on a weekly basis (this also depends on the size of the country and the market share of the mobile operator. In Austria it can be hundreds of GB per week but in Germany it may be 10 times more). The kind of data storage and delivery depends on the mobile service provider. Some providers deliver all logged technical information in a customised binary or proprietary file format and the investigating teams can extract the necessary geographic coordinates and time stamp information from sequential log entries by writing special programming scripts. Other providers deliver only pre-processed or even aggregated data—e.g., numbers of active subscribers in a mobile network cell occupying a raster cell during certain time steps. In this case no motion information can be extracted from the data. Motion information can only be extracted if (an anonymised) device ID is provided with the location and time stamp information to track single trips. Otherwise only user distribution or “user density” patterns can be depicted. Figure 2 depicts a sample CDR in machine-readable format.

Sample data stream (Hexadecimal)

ABD00012DEFF304205089076002020404002AAAABBBBF0ADEF0000ECC0001135

**Figure 2.** A sample mobile CDR example (hexadecimal format).

Motion exploration is heavily dependent on mobile phone and auxiliary city data, as indicated in Tables 3 and 4. Tables 3 and 4 also indicate what information can be generated from different datasets and which datasets have been used in Motion Explorer (Table 3). This approach enables city administrations to make a close approximation of the overall effort/cost for acquisition and processing of data for Mobility Explorer in different cities.

Since the CDR data are stored and provided in files of varying sizes, for pre-processing of CDR data the complexity can be $O(n^2)$. This means each anonymous ID is compared with other occurrences of the same ID in the file. This means a higher number of transactions will slow down the processing time. This complexity further increases when processed CDR is correlated with land-use data. This will double the complexity, i.e., $O(n^2)^2$. Different optimization techniques can be applied, e.g., map-reduce by using Hadoop to divide larger files into smaller size files and process them on Hadoop worker nodes. However, synthesizing results for the same user across multiple files will require careful algorithms to provide uniform mobility patterns for selected users and correlate with geographical points of land-use data.

NB: In comparison to other research projects (cf., e.g., [26]) the data provided for the development of Mobility Explorer contained mobile device movement data for single days only, since, for privacy reasons, the Austrian provider of the CDR data is switching the anonymous IDs each day at 12:00 p.m., so pattern analyses over whole weeks or months (like in other projects) are not possible with these kinds of datasets, since mobile users of one day cannot be identified after 12:00 p.m. For this reason, Mobility Explorer is focusing on dynamic visualisations of movements of bigger groups of mobile phone users, e.g., commuters to let, e.g., city administrations check for newly arising streams of commuters that have developed only recently and have not been detectable in the past since the tools to detect this have been missing.

Table 3. Datasets used within the Mobility Explorer (ME) application.

Dataset Title	Usage of Dataset in Mobility Explorer	Dataset Type	Data Type				Data Source
			Vector	Raster	Tables	Raw	
Basemaps	Basemap for orientation purposes	Quarter level	X		X		Planning authorities
	Basemap for orientation purposes	City level	X		X		Planning authorities
	Background layer for selecting analysis region	Census units (at least 1 per km ² raster cell)	X		X		Planning authorities National statistics
Urban Topography	Basemap for orientation and visual analysis purposes	Satellite/ortho-images		X			Planning authorities
Population (Census)	Basis for statistical projection of mobile phone users to total population	Population Total	X		X		National statistics
Mobile Phone Data	Basis for visual analysis and retrieving mobility analysis data	Raw data stream				X	Service providers
	Basis for visual analysis within ME	Pre-processed data			X		Own calculations
	Basemap for orientation and visual analysis purposes (only available in Vitoria-Gasteiz application)	Cellular antenna locations	X				Service providers,
	Basis for statistical projection of mobile phone users to total population	Mobile Phone Total			X		National statistics

Table 4. Datasets to be considered for future applications.

Dataset Title	Usage to Consider in the Future	Dataset Type	Data Type				Data Source
			Vector	Raster	Tables	Raw	
Basemaps	Basemap for orientation purposes	Block to block clusters	X		X		Planning authorities
Urban Topography	Basemap for orientation purposes	Transportation network (road network)	X				Planning authorities
	Basemap for orientation and visual analysis purposes	Public transport network (railways, light rails, trams, sub-ways, bus routes, stations)	X		X		Local transport authorities
	If available, useful for bias detection within mobile phone datasets.	Mobile phones by age classes (0–15 years, 15–45 years, 45–65 years, 65+ years)			X		National statistics
Population Data (Census)	If available, useful for social bias detection within mobile phone datasets.	Population by gender, by social class characteristics by education			X		National statistics
	If available, useful for detection analyses of roaming/temporary available users.	Temporary population (students, tourists, etc.)			X		National statistics
	Basemap for orientation and visual analysis purposes.	Population by offices, education, social services, sports, shopping, spare time	X				Local authorities
Households	Basis for statistical projection of mobile phone.	Households Total			X		National statistics
Workplaces	Basis for statistical projection of mobile phone users.	Workplace Total			X		National statistics

5. Case Study: Vienna—Stakeholders & Requirements Element

One of the main objectives of transport planning in the city is to keep the distance short and shift from private motorised traffic mode to non-motorised transport or public transport. However, the existing data on the use of public spaces, population distribution and mobility behaviour (e.g., poll data or traffic counters) have some limitations (see Table 1). Therefore, the City of Vienna's main goals here are to derive information from additional sources about diurnal population distribution and real mobility behaviour of citizens that can be used to: (i) compare and contrast existing polls, statistics and modelling results and in certain cases complement existing information; (ii) get better insights in mobility and traffic behaviour of the citizens (especially origins and destinations of trips); (iii) get information about the attractiveness of selected areas; and (iv) use this as evidence to improve transport and urban planning initiatives in the city. This information permits the targeting of areas of attraction within the city, the opportunity to obtain a full understanding of the basis for this attraction, and accordingly to create urban and transport planning responses. Hence, the City of Vienna has a great interest in investigating the potential of using mobile phone data to visualise and analyse diurnal population distribution and public motion dynamics to achieve the above goals.

Vienna considers mobile phone data an entirely new source of information concerning socioeconomic and mobility phenomena. Specific information offered includes population density at various locations at various points of time to identify usage and attractiveness of specific places. Furthermore, travel behaviour analysis by exploration of daytime traffic flow (e.g., Origin-Destination (O-D)-matrices) provides better insight into diurnal public mobility and traffic behaviour in the city at different hours of the day. Finally, the mobile phone data reflect the real behaviour of a very large sample in a high spatial and temporal resolution at reasonable costs. Mobile phone data are appreciated as an extension to existing polling and counting data.

In general, Vienna needs to get an idea about the population distribution patterns within the city during workdays and weekends, i.e., how many people are staying in which district or area. Vienna is interested in detailed information about occupancy patterns in the densely built-up areas during day and night hours as well as in recreation areas during weekend hours. Vienna needs population density data for average work days (map and table) for 0:00, 5:00, 6:00, 7:00, 8:00, 9:00, 10:00, 12:00, 14:00, 15:00, 16:00, 17:00, 18:00, 19:00, 20:00 and 22:00, and at the weekend for 10:00 and 15:00.

Vienna is also interested in a visual interactive tool whereby temporal and spatial aggregates like census districts can be shown on a map. Furthermore, Vienna is interested in Origin-Destination (O-D) matrices between districts. Also, Vienna needs intra-city and extra-city origin-destination matrices containing the total traffic. The question Vienna wants to answer is: Where do people live (e.g., sleep = 0:00) and where do they work (e.g., 10:00)? This is especially important for designing the transport system. At the weekend the same question is important for improving the attractiveness and accessibility of public and/or open spaces. The spatial resolution of the provided data is 1000 m grid cells. The City of Vienna is aware that these data reflect only a subset of the total traffic, but the data can be used to support existing population distribution and mobility datasets.

6. Mobility Explorer System Architecture—Standards & Tools Element

In order to respond to Vienna's requirements and explore CDR with additional city data, Figure 3 depicts the Mobility Explorer system architecture. The Mobility Explorer is designed as a web application consisting of a client side interactive map interface using the GeoExtJS and OpenLayers libraries (the underlying geo processing and visualisation can also be performed using commercial tools like ArcGIS (subject to licensing fee) or open source tool like QGIS, Geoserver and specific JS libraries). The server side consists of a PostgreSQL/PostGIS database. Mobility Explorer features and a Tomcat servlet as well as a Geoserver allow us to visualise the pre-processed motion data derived from the mobile phone logs. These tools are commonly used in geo-spatial application development.

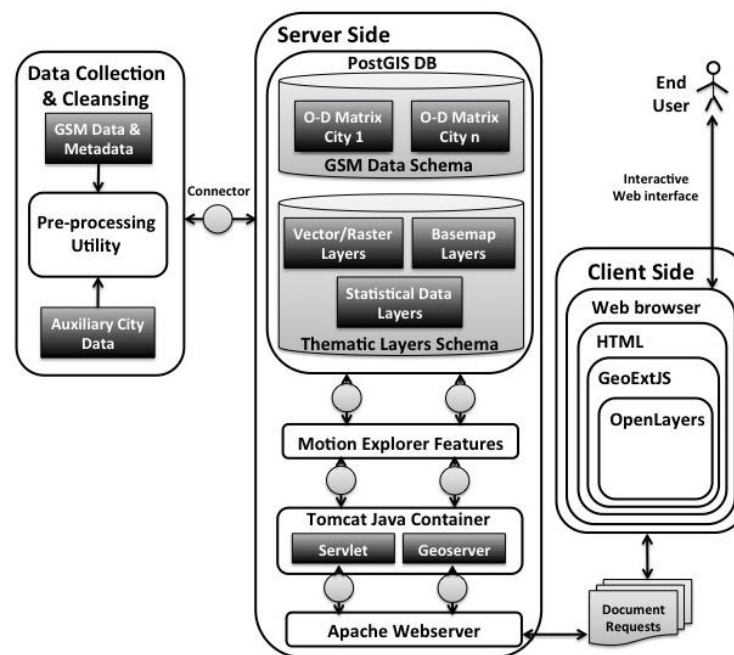


Figure 3. System architecture of the Mobility Explorer.

The Mobility Explorer derives and visualises specific features requested by end users, e.g., heat maps, extrapolation of mobile phone data to overall census population, etc. The data collection and cleansing component is mainly used to pre-process mobile phone data and make it error-free. It can also be used to process additional city data, e.g., statistical data, basemaps, etc., before storing them in the PostGIS database.

In order to have a flexible structure to add mobile phone data from different providers in countries (and cities), the mobile phone data has to be pre-processed (for error detection and exclusion—see next section) and stored in a database, as depicted in Figure 3. For experiments, Vienna city-region CDR data for a sample week in year 2012 were acquired from a major mobile communication service provider. This allowed for performing a flexible exploration of diurnal subscriber distribution dynamics over time, aggregation of subscriber distribution to individual spatial entities and interaction and motion pattern analysis by aggregating single trips during the day to be extracted based on (anonymous) ID and time stamp information. The process steps to derive the Mobility Explorer visualisation dataset from the raw CDR data are shown in Figure 4.

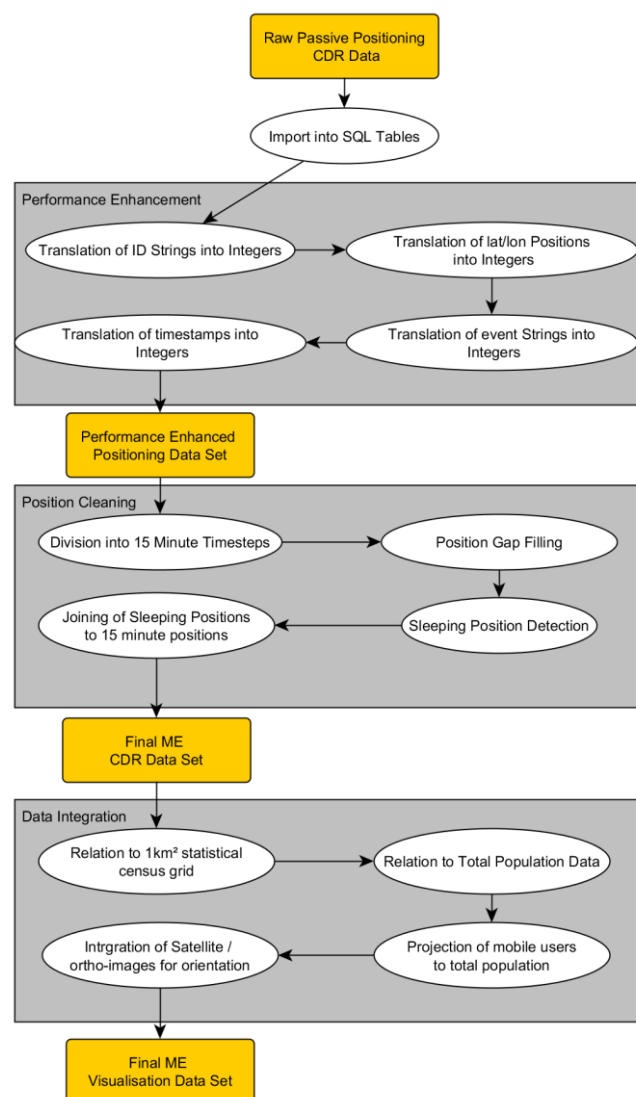


Figure 4. Processing steps: from raw CDR data to the dynamic visualisation dataset.

7. Motion Exploration: Vienna Application Outcomes—Outcomes & Evaluation Element

7.1. Data Aggregation, Processing Optimisation and Accuracy Experiment

One of the methods to map mobile phone user distribution and visualise on city maps is to aggregate all customers observed within certain areas or at one location in a certain time range. These areas or locations are either the cell towers (i.e., representing districts, zones, network cells or grid cells) or triangulated positions based on cell tower locations and signal quality. Any spatial entity can be used to depict patterns. The smaller the network cells and the smaller the analysis entities (census districts, traffic cells) the finer is the distribution pattern. In order to optimise the CDR processing, the data must be sorted by time stamp and grouped by same (anonymous) IDs for visualisation of results. Also, all double counts of IDs for the observed time range must be erased during the data pre-processing stage. In addition, the anonymous IDs should be replaced with integer numbers for faster processing (see also Figure 4, top grey box (*Performance Enhancement*)).

The above approach has been applied on Vienna CDR to depict the diurnal population distribution dynamics. However, during the experiments and development of the Mobility Explorer, several data problems have been encountered, mainly because of the nature of the mobile phone data. These problems were:

- *Pre-processing CDR*: Data pre-processing is needed to understand binary mobile phone log files. This pre-processing revealed quality and reliability issues in the mobile phone data by visualising it through maps. The first test dataset proved to be erroneous as it turned out that some of the cell towers in the West of Austria reported false positions. A similar issue was detected in the Vitoria-Gasteiz dataset due to its low density of cell antennas (see chapter 8). Such errors—if not detected—can lead city planners to make false assumptions about mobility patterns and population distribution. So, a detailed exploration of the mobile phone data regarding errors and quality before applying the data is absolutely essential.
- *Optimisation of data logger*: Typically the mobile phone data logger stores all log records in a sequential manner without indexing. This needs to be optimised for large number of records. As the dataset has a very high temporal resolution because of technical monitoring reasons, this temporal resolution has to be reduced to avoid redundant location information (to accommodate the high volume, velocity and variation of data, the common data model could, e.g., be in JSON format and stored in NoSQL (e.g., MongoDB) in the future). For the Vienna application, a 15-min time step was selected to explore and view temporal dynamics. This 15-min time step was acceptable for planning needs. The analysis of the data logger revealed that the mobile phone log data does not provide a unique motion pattern of devices. Therefore, extra processing was needed to identify and build a chain of trip positions of unique devices.

However, during the chaining process further position errors were detected, i.e., mobile devices were observed to be “jumping” to far-distant locations within the 15-min time steps. It turned out that these pseudo-movements occur due to load balancing reasons, i.e., heavily used cell towers redirect devices that exceed the cell tower’s carrying capacity to other cell towers with less data traffic. This sometimes made it difficult to identify the valid location of mobile devices. Such location errors either need to be corrected or removed from the dataset. Usually those new positions that would require impossible travel speed to reach the cell within this time were re-mapped to the last valid position to avoid wrong trips. This technique was applied to the whole log data file and then a small experiment was conducted to check if the log data are still able to deliver correct tracking information for a device. For this test, the GPS and the mobile phone position—based on the cell tower positions of one anonymous device during a trip (Figure 5)—were compared. The test proves that the location corrections in the mobile phone log data are delivering appropriate movement information (over one day).

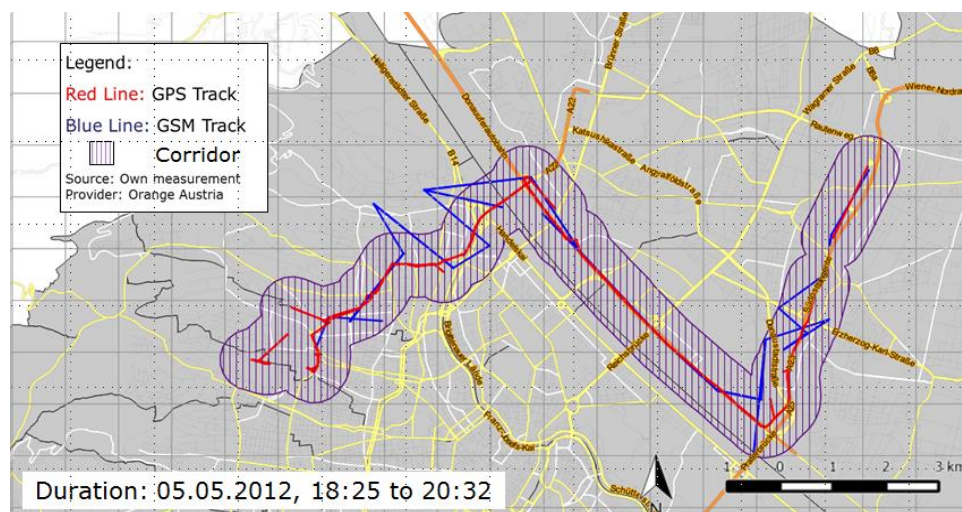


Figure 5. Comparison between a GPS track (red) and the corresponding cell tower locations (blue) of a trip of a test user from the 22nd district to the 18th district in Vienna. The purple striped area depicts the accuracy corridor of 1 km.

7.2. Visual, Interactive and Dynamic Cell Occupancy Analysis—City Scale

The processed CDR has been used to visually represent population distribution and mobility dynamics as per Vienna requirements (Section 5). One important question is whether CDR is representative enough as compared to actual population and in this respect the following analysis was performed.

- Population representation: CDR vs. population census:** Since the CDR data in the Vienna case have been coming from one single provider, there was also a need to check the feasibility of projecting the mobile phone user data to the whole city population. This was done by comparing the 1 km² grid cell inhabitants of the national statistical institute with the aggregated 1 km² grid mobile phone user numbers for the time period of 0–6:00 (of one single day), which was considered to be a (rough) representation of the “sleeping population” of Vienna (this approach has been taken because it was considered that users/devices during this period of time would—in most cases—reside in the places where the users are living (i.e., “sleeping”, with the exception of users working at night, e.g., policemen, taxi drivers, clinic staff, etc.)). The map in Figure 6 shows the result: The statistical number of inhabitants per 1-km² cell (depicted as coloured “dots” in the map) matches the spatial distribution of the mobile devices (depicted as coloured “rectangles”): Where there is a high number of inhabitants there is also a high number of mobile devices; where there is a low number of inhabitants there is also a low number of mobile devices.

Figure 7 depicts the map’s results in a statistical graph showing the relation of statistical 1 km² population to 1 km² “sleeping population” from the mobile phone user dataset: The value of $R^2 = 0.5533$ shows quite a good correlation between the observed mobile devices and the statistical population in the 1 km² grid cells, which leads to the assumption that it is indeed feasible to calculate a coefficient for each source cell:

$$coefficient_{source_cell_x} = \frac{\text{statistical inhabitants}_{source_cell_x}}{\text{sleeping population}_{source_cell_x}} \quad (1)$$

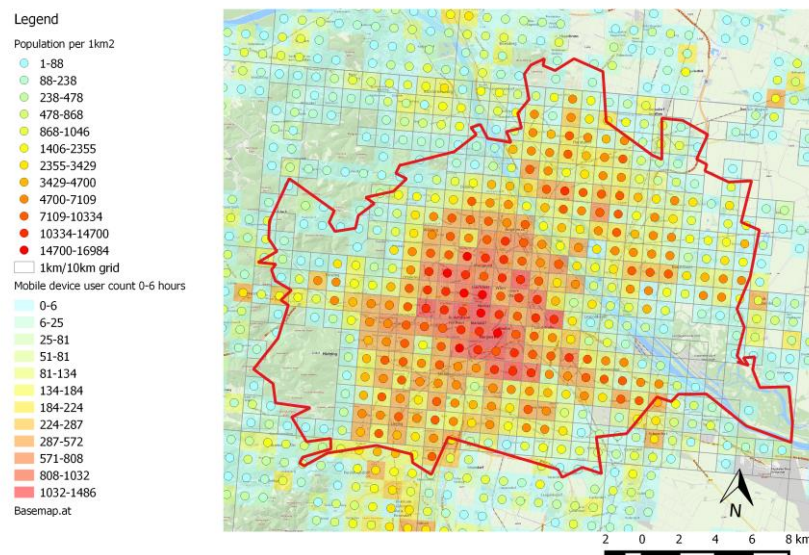


Figure 6. Visual comparison between 1 km² statistical population of Vienna (dot signature) vs. 1 km² mobile phone user densities between 0 and 6:00 (square signature).

When a source cell is selected in Mobility Explorer, each number of users spreading over time over the city (i.e., over the raster cells within the map) is multiplied by this coefficient to calculate the representative statistical population coming from the source cell.

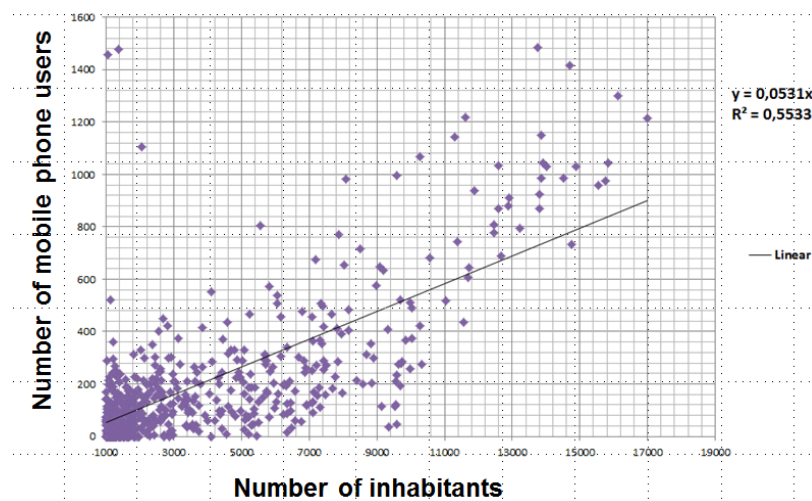


Figure 7. Statistical comparison between 1 km² statistical population of Vienna vs. 1 km² mobile phone user densities between 0:00 and 6:00 ($R^2 = 0.5533$).

One unique feature, which we were unable to find in any other related work, is providing an interactive visual map that allows for performing dynamic cell occupation and mobility analysis at the city scale:

- **Dynamic mapping of diurnal motion dynamics for interactively selected cells (origins):** This functionality lets the user choose any “source cell” (a red rectangle in the map) that is defined as the starting cell of the identified “sleeping population” of that cell. Using the time slider, a user is then able to see the movement of this group of users (originating from the chosen cell) through the city/region over the day (displayed as a heatmap of user densities—Figure 8).
- **Presentation of the diurnal occupancy of cell “visitors” targeting an interactively selected cell:** By clicking on one of the heat map raster cells (in Figure 9, below the cyan rectangle), it is possible to get an overview of the diurnal densities of users (coming from the “source cell” chosen in the first step (red rectangle)) during the day, which is displayed as a line graph at the bottom of the application (Figure 9).

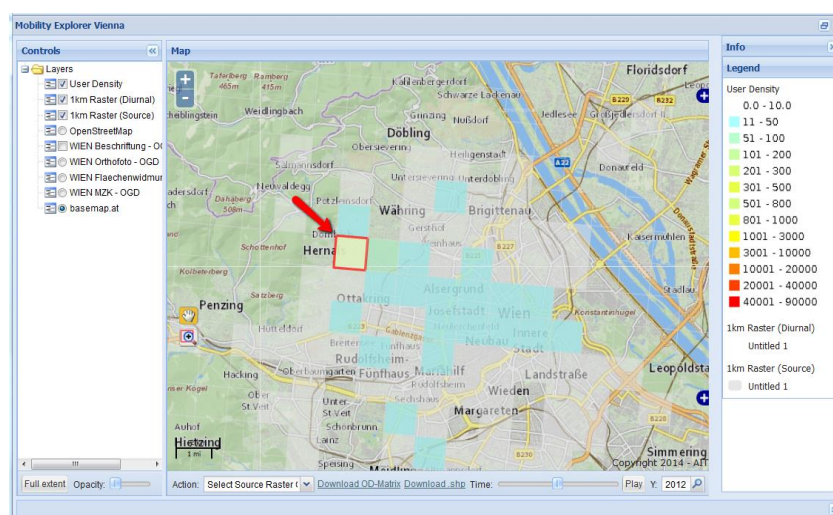


Figure 8. Dynamic mapping of motion dynamics for interactively selected source cell (selected cell marked by bold red outline).

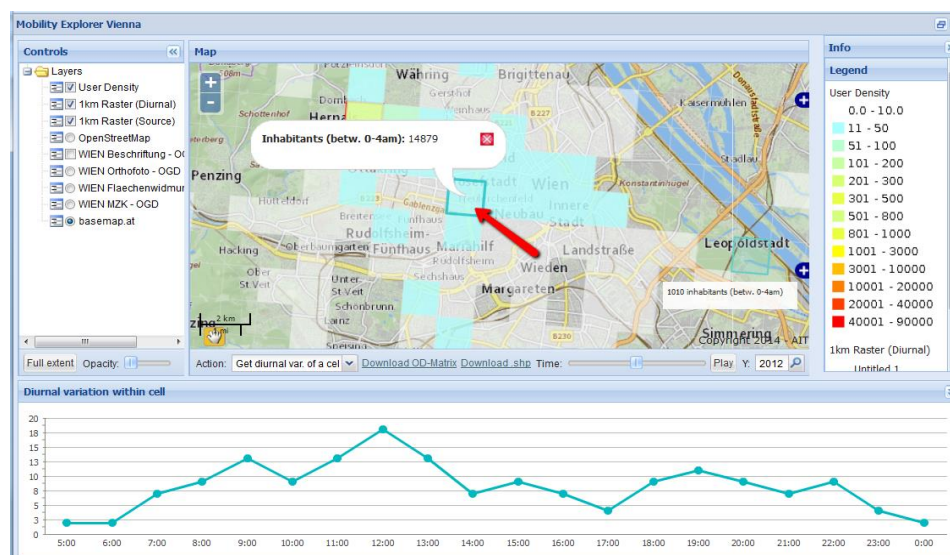


Figure 9. Map accompanied by a line chart of the diurnal cell occupancy of visitors of an interactively selected target cell in Vienna (cyan outline) from selected source cell (red outline).

Both Figures 8 and 9 demonstrate the temporal cell occupancy between two selected cells in the city. As an example, Figure 10 depicts the difference in population distribution according to two different target cells selected on the map (timestamp: 14:30). The left-hand figure shows that the population from the chosen source cell has spread to a much wider area of the city of Vienna than the one from the source cell in the right-hand figure.

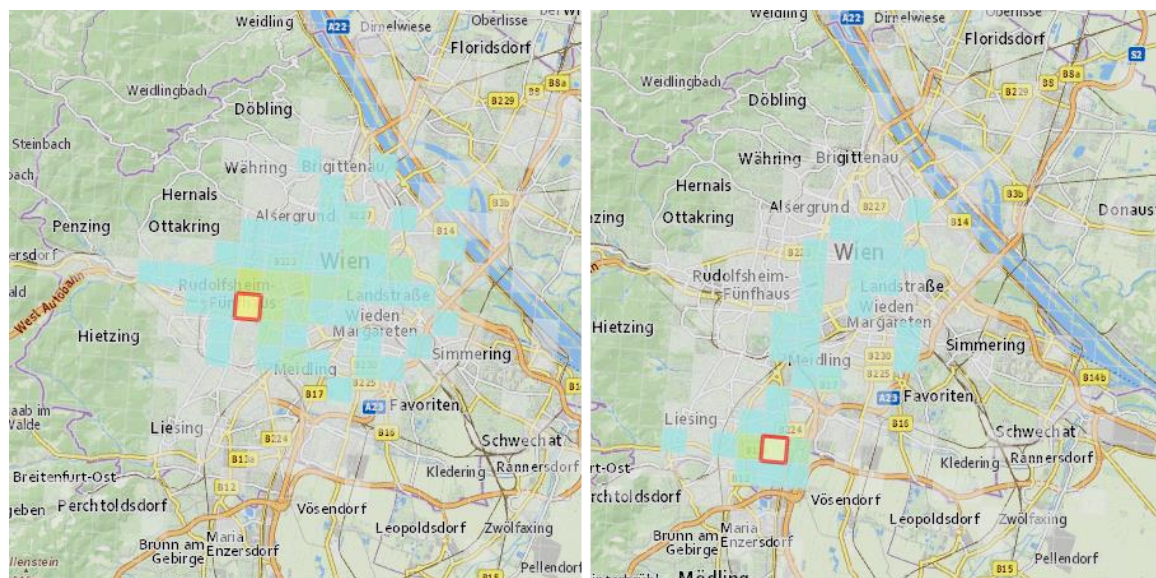
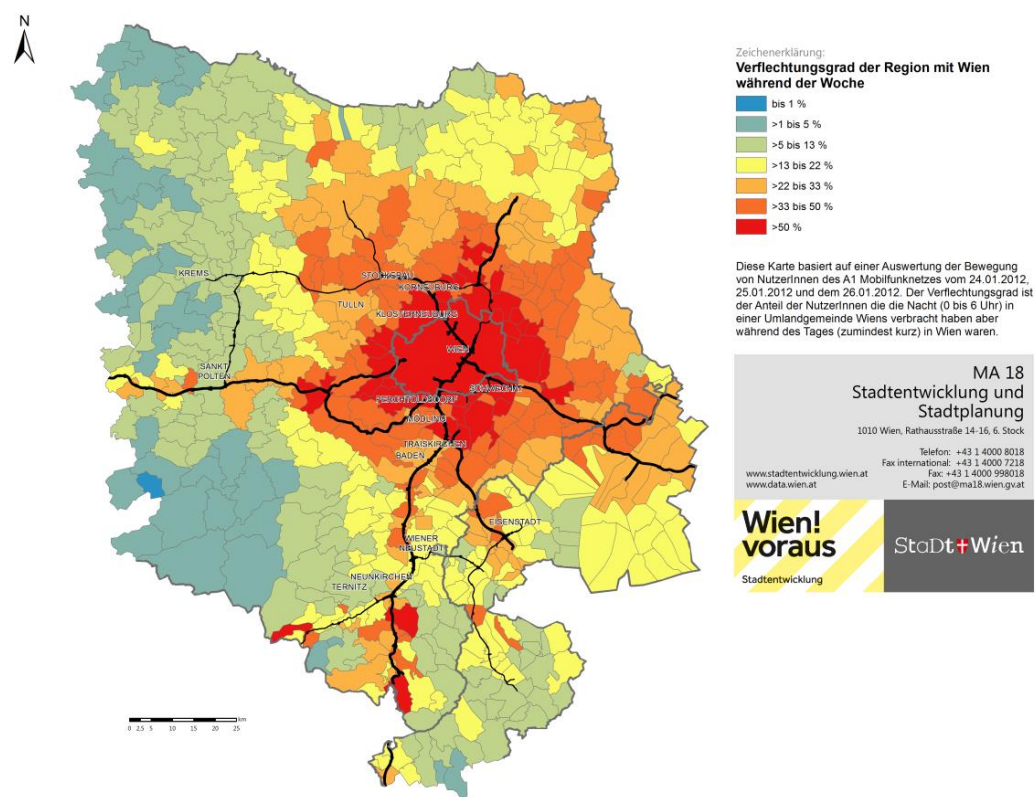


Figure 10. Comparison of target traffic from two different source cells (left versus right) at 14:30; source cell: red rectangle.

The above interactive and dynamic population distribution and mobility analysis at the city scale can also be saved as origin destination matrices (example in Appendix A) that provide new information source to planners that is not possible to collect using conventional methods. Planners and GIS experts can integrate this information with other data sources for further analysis. This information is also useful for analysing population dynamics at the city-region scale to determine functional urban areas

and identify mobility patterns to perform evidence-based analysis of city policies, e.g., functional urban region, city of short paths, etc.

- City-region scale—regional interdependency:** Mobile phone location data allow Vienna to map spatial interactions between the city-region and the City of Vienna. Figure 11 depicts regional interaction patterns between the Greater Vienna region residents and the city. The analysis of motion exploration data reveals that values are significantly higher than in classical commuting maps based on census data. This is due to the fact that mobile phone data include all trips (work, leisure, shopping, education, etc.) while Austrian census data include trips to work only. Looking at the high percentages depicted in the legend, we recognise how remarkably interwoven the Vienna Region is. So this comprehensive dataset can substantially contribute to identifying the functional urban region—a research question often discussed among regional planners. Another aspect is the distinct impact of high level transport infrastructure on people’s mobility behaviour. This applies to motorways (black lines in the map) as well as railways (not depicted in this map). Figure 11 reveals that the level of interdependency along these axes is especially high.
- City of Short Paths:** One important goal of the transport policy of the City of Vienna is a reduction of the average length of trips and thus the total kilometres travelled. This policy can have a huge impact towards achieving Green House Gas (GHG) emission targets and improving the quality of life of the general public. However, it is difficult to collect holistic evidence to assess such a policy and/or initiate new plans. In this respect, the mobile phone data exploration reveals evidence-based results, as depicted in the two maps below (Figures 12 and 13). Figure 12 shows that people living in the inner districts travel fewer kilometres, which indicates that the majority of people are not taking long journeys for their routine activities.



Regionale Verflechtung

Figure 11. City-region regional interaction pattern mapping using Mobility Explorer (German labels are used to support local language).

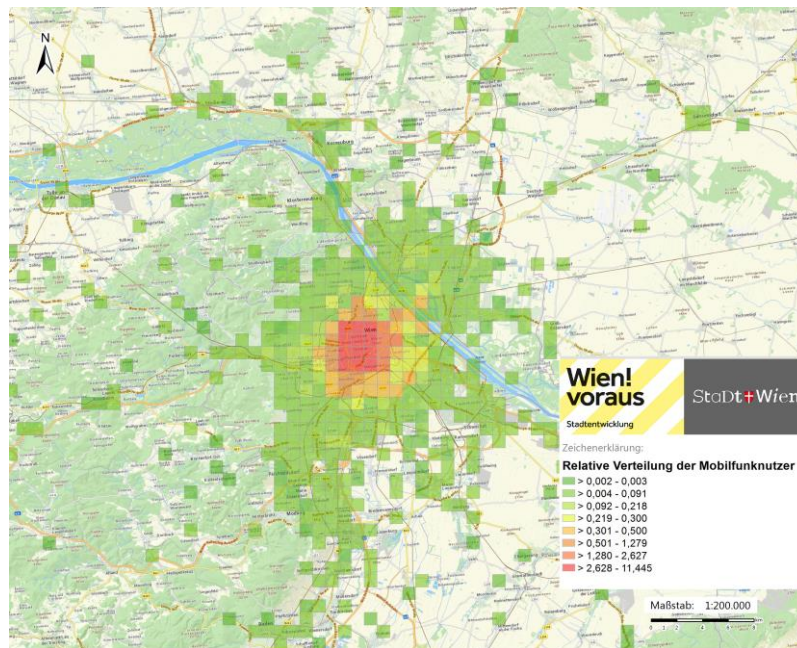


Figure 12. Mobility pattern of residents living in inner districts of the City of Vienna (German labels used to support local language).

In contrast to Figure 12, Figure 13 shows that people who live in suburban areas have to travel to inner districts for routine activities. For instance, people from southern suburbs have a high percentage of travel as compared to suburbs in the north, east or west of the City of Vienna. Such information at such a large scale provides a stimulus to initiate new planning projects including creating job opportunities near the southern suburbs so that people will not have to travel far.

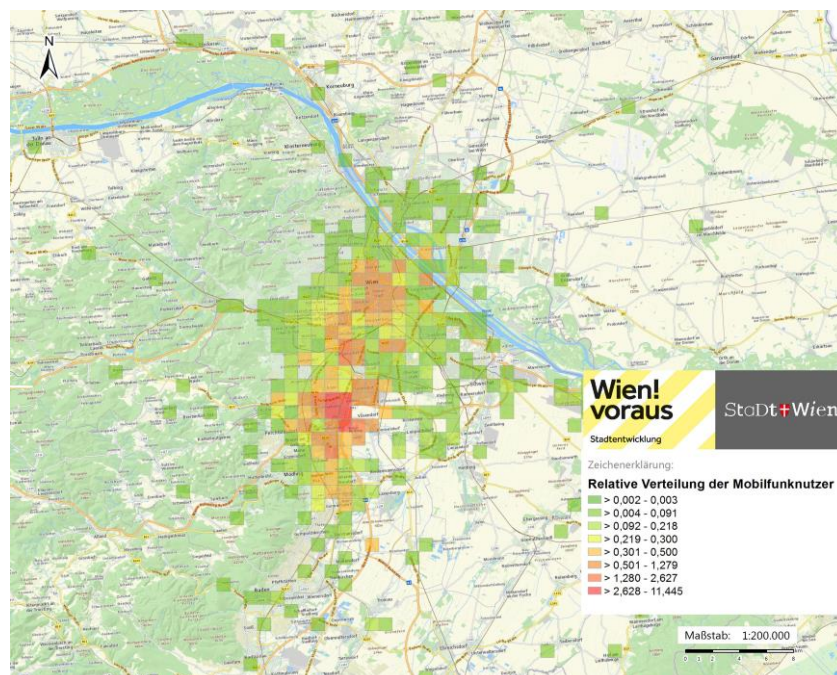


Figure 13. Mobility pattern of suburban residents living in the south of Vienna (German labels are used to support local language).

7.3. External Evaluation Results

In addition to the above results of Vienna application, the motion exploration has also been evaluated by domain experts and city representatives from Vitoria-Gasteiz, Bologna and an external advisory group representing Pan-European organisations dealing with urban, mobility and city governance domains. The Criteria Indicators and Metrics approach was applied to perform the evaluation [27]. The overall evaluation participation rate was very promising as 16 expert users with different roles and expertise, i.e., urban planners, policy makers, GIS experts, IT experts and others, participated in the evaluation of the motion exploration tool. This evaluation was performed with the objective of assessing the usability, functionality, benefits and relevance of the motion exploration tool in collecting the evidence base for smart city planning. The evaluation results clearly indicate the benefits and relevance of the application in achieving research objectives such as support to urban and transport planning and decision-making. In the above sections, the Vienna results presented the benefits and functionality of motion exploration; due to space limitations, we briefly present the usability evaluation results here.

Most of the evaluators found the application easy to understand and use but also indicated that the resolution of maps could be enhanced and intuitiveness can be further improved, e.g., by providing context-sensitive help. The tool allows end users to interact with the city map and explore population distribution and mobility patterns. It also allows for running simulations showing dynamically changing space occupation for various zones of the selected city. In terms of overall performance, 80% of evaluators agreed that the response time to user queries was reasonable but 20% disagreed. This is mainly due to the slow response from the remote server machine where motion exploration was deployed. This suggests that, due to high processing demand, motion exploration should be deployed on a powerful server dedicated to this application so that the response time to multiple users' queries is acceptable. All evaluators disagreed that the system is complex. This may be the reason why 40% to 60% of evaluators agreed that they would like to use this system frequently but 40% of responses were neutral, which may be related to the specific needs for this system in their organisations' core businesses.

Up to 60% of evaluators agreed and 40% were neutral that different features are well integrated and accessible by one GUI. This neutral response can be referred to the need to synchronise GUI elements (active/de-active) when a specific operation is performed. For example, while the simulation is running to show the population distribution at different hours of the day, other features are still accessible and can be activated to interrupt the simulation.

Up to 80% disagreed that the system is cumbersome to use; the 20% neutral response may be attributed to the need for intuitiveness and easy-to-use functions using application GUI. About 80% of evaluators disagreed that they had to learn too many things before they could get going with the system, but 20% agreed. These results suggest that the system is not too difficult to learn and use as it has well-integrated and consistent functionality accessible through its GUI, but there may be additional training needs depending on the skillset of end users and there is potential to improve the application GUI towards further intuitiveness and usability.

Overall, the evaluators found Mobility Explorer easy to use to produce required outcomes (e.g., Origin-Destination (O-D) matrices, high-resolution images, etc.) that can be further utilised in urban and transport planning processes. Likewise, positive responses were obtained for functionality and benefits-related evaluation criteria.

8. Lessons Learned and Recommendations

As presented in the previous sections, the Mobility Explorer framework provides all necessary elements to develop a visual tool using data from a variety of sources that can satisfy the real needs of city stakeholders. In this process of applying the framework to the City of Vienna case study, the following main lessons have been learned:

1. The spatial accuracy of the mobile phone data needs to be considered critically when it comes to the development of motion exploration applications for a city. It was not until the CDR data had been visualised and thoroughly checked that the participating cities realised certain questions they wanted to be answered in their application requirements could not be answered due to the coarse granularity of the data. For instance, the mobile phone data for Vienna proved to be accurate (e.g., Figure 5) but in Vitoria-Gasteiz, due to its relatively small geographical urban area and the low density of mobile phone antennas, it was difficult to study the movements of the cell phones accurately. Also, some of the mobile phone antennas were positioned on a mountain rim, leading to a lot of connections due to their good “visibility”. Figure 14 shows the low density of Vitoria-Gasteiz’s antennas. This less dense spread of antennas in Vitoria-Gasteiz resulted in a lot of inaccurate location information for mobile devices and hence proved to be less usable for motion exploration.

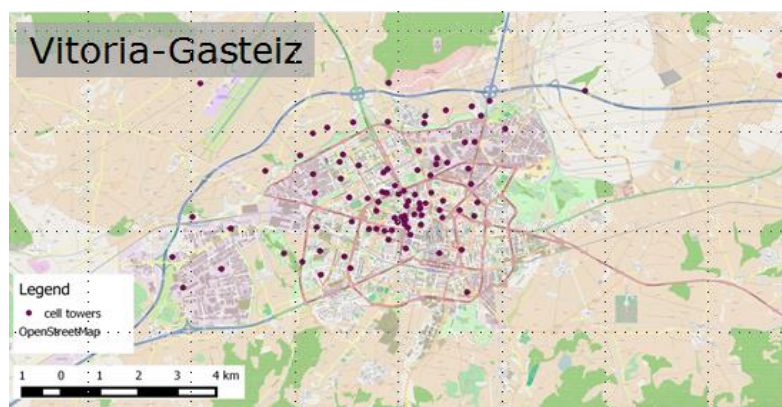


Figure 14. Coarse density of mobile phone cell tower locations in Vitoria-Gasteiz.

2. The different datasets of the national providers differed hugely in their temporal and spatial resolution, leading to limitations when trying to extract motion information. To overcome this limitation and the heterogeneity between different mobile phone datasets (from different providers), a data standard for mobile-phone-based log data needs to be defined for wider adoption by cities and businesses. This can be further facilitated by designing a common data model with data harmonisation and integration guidelines to enable cities to use CDR from different service providers for motion exploration.
3. The level of detail in CDR vary from one data provider to another, which in specific cases may limit the extent to which rigorous mobility analyses can be performed. For instance, using the above mobility analysis techniques it is difficult to determine travel mode from CDRs, e.g., walking, cycling, car, tram, bus, etc.
4. Mobile data privacy is a critical issue. All companies that provide CDR data (for example, in the case of Mobility Explorer from Austria, Italy and Spain) have strict policies about sharing such data, e.g., some provide aggregated numbers that do not cause any privacy concerns. In some countries, any personal data that can be linked to individuals is not allowed to leave the country. Others deliver anonymised raw data, where single movements of devices could be identified to some extent, although the ownership of the device and thus the individuals’ privacy is respected and the data are fully anonymised. Many companies do not offer data provision service and the data are delivered on special request and only to selected customers. CDR is secured in log files by not storing individual subscriber information in the log. Renewing the anonymous random user ID on a daily basis hinders long-term observations of a single entity to protect privacy. Further details can be found in IMSI [28]. While visualising or generating Origin-Destination matrices, it must be ensured that data privacy is protected by applying techniques like cell aggregation, line

trimming from source and destination, etc. Also, care must be taken when working with these datasets to comply with privacy and data protection guidelines, e.g., data of single individuals must not be depicted, collective behaviour patterns must avoid identifying individuals, location information shall be fuzzified to hinder the identification of single positions, etc.

9. Conclusions and Future Research Directions

The impact of CDR in smart city planning applications goes beyond the fields of academia and research. The proposed Mobility Explorer framework, operationalised through the development of the Mobility Explorer for the Vienna case study, has provided detailed insight into CDR processing, data quality and integration, and generated new knowledge that otherwise is difficult or expensive for city administrations to acquire. In this regard, the Mobility Explorer focused on the following unique features that have not been provided by any other similar application:

- Visual interactive maps that allow end users to dynamically map diurnal motion dynamics for selected cells;
- Presentation of the diurnal occupancy of cell “visitors” targeting an interactively selected cell;
- Testing the application of CDR against city planning requirements by engaging with domain experts (urban and transport planners);
- New visual outputs, e.g., heatmaps, accuracy of CDR maps, etc.

In addition to the above, the Vienna application reveals a number of benefits of the Mobility Explorer and its usefulness as a new information source to city administration that can result in evidence-based planning and decision-making. This application also demonstrates that the processed information provides a new evidence-based source that is either not available through conventional data collection methods or is too expensive for city administrations to collect at a city or city-region scale. The experiments also revealed data processing challenges and recommended techniques to deal with data quality issues such as location accuracy. Also, these experiments demonstrated that the high or low cell density or technical infrastructure of mobile service providers play a crucial role in determining the effectiveness of CDR in visualising diurnal population distribution and mobility dynamics.

The Mobility Explorer evaluation results are very positive in the sense that they indicate strengths and identify potential future directions of development. The functional appropriateness is validated at the same time that limitations are identified. Most of the results indicate that the tool is an effective source of new information to identify population distribution and mobility patterns across a city (or at the city-region scale) that can potentially be used for urban and transport planning and policy making. This proves the research hypothesis that *‘mobile phone data can be used in deriving and visualising diurnal population distribution dynamics and sojourn mobility patterns to derive necessary information intelligence for smart city planning processes’* to a certain extent as there were also some limitations identified in mobile phone data. Among these limitations are the lack of appropriate detail in the available CDR, which makes it difficult to fulfil all the requirements of planning users, e.g., social biases, travel mode, live/active and dead phone connections. Nevertheless, the application itself has the potential to contribute significantly in terms of gaining insight into space usage and mobility analysis. The overall population distribution and mobility patterns represent an approximation but are useful to gain this kind of new information which otherwise is not available to city planners or too expensive to acquire at city or city-region scale.

Detailed analysis of the Mobility Explorer applications has identified future research directions including real-time observation and traffic control if such data can be provided as streaming data by mobile data providers. This would also require designing a common data model for Mobility Explorer and defining guidelines for data harmonisation and integration from different mobile service providers that will enable a large number of cities to perform similar analyses. This would also require support of dynamically scaling systems like cloud computing to accommodate processing and storage demands at peak times. Another interesting research angle would be to combine mobile phone data

with social network information to increase the reliability and information density of the resulting combined datasets [19].

Acknowledgments: The research work presented in this paper was partly funded by the European Commission's Framework Programme Seven Project UrbanAPI.

Author Contributions: Jan Peters-Anders and Arno Breinbauer conducted the pre-processing and analysis of the raw mobile phone data and programmed the database routines to do so. Jan Peters-Anders furthermore programmed the web application that visualises the movement of the population. Zaheer Khan and Wolfgang Loibl worked on the requirements for the application of mobile phone data movements within the city context. Zaheer Khan also contributed to mobility data framework, architectural illustration, evaluation and discussion chapters. Helmut Augustin was responsible for testing the results of the analysis within the city administration of Vienna.

Conflicts of Interest: The founding sponsors had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, and in the decision to publish the results.

Appendix A

Table A1. Matrix of references dealing with the topic CDR data analysis and visualisation.

Ref. No.	O-D Matrix		Temporal and Spatial Population Distribution			Temporal and Spatial Mobility Patterns			Data Issues	Information Overlays			Pre-Processing	Interactive Maps	2D Maps	Privacy
	District-to-District	Point-to-Point	Weekdays (or Not Disclosed)	Weekends	15 Min	Cell-to-Cell (or General Mobility)	City-Region	Mode Detection		Census	Demographics	Gender Analysis	e.g., Cleaning for Noisy Data or Accidental Edges			
[19]		x	x			x			x	x			x		x	
[26]	x	x				x	x			x					-	
[29]															x	x
[30]	x		x		x				x	x			x			x
[31]			x	x		x	x		x				x			
[32]	x	x					x	-	x				-	x	x	x
[33]	x	x	x	x	-	x	x	x	x	x	x		x	-	x	x
[34]	x		x										x		x	x
[35]	x	x				x							x		x	x
[36]	x	x	x	x	-	x	x			x	x				x	
[37]	x		x			x			x				x		x	x
[38]			x			x			x						x	x
[39]	-		-			-			x	-			x		x	x
[40]	x	x	x	x	30 min	x	x		x						x	
[41]			x	x		x	x		x				x			
Mobility Explorer	x	x	x		x	x	x		x	x			x	x	x	x

Table A2. Summary of related work in comparison to Mobility Explorer (cp. also Table A1 above).

Ref. Nr.	Title	Comparison to Mobility Explorer
[19]	* A survey of results on mobile phone datasets analysis	Provides a survey of CDR used for exploratory social network analysis or mobile communities formed based on call records. This helped the authors in comparing observation data against self-reported surveys and has resulted in concluding that self-reported surveys produce subjective bias and vary significantly from the reality. The authors recognise that dynamic or temporal dimension of data analysis is rather recent area of research. It appears that the application of UrbanAPI is (still) unique from the perspective of establishing dynamic mobility patterns, interactive visualisation and integration with land-use data to satisfy real city planning needs.
[26]	Human Mobility Modeling at Metropolitan Scales (2012)	This paper discusses a procedure called WHERE (“Work and Home Extracted REgions”) It describes a model to calculate probability distributions.
[29]	Space, time and visual analytics (2010)	Not relevant in the GSM context of this Mobility Explorer paper, but highly relevant because of discussion of visualisations, especially concerning interactive map applications (cp, Chapter 4.2)
[30]	* Urban Sensing Using Mobile Phone Network Data: A Survey of Research	A highly relevant survey paper. It outlines CDR and discusses strengths and weaknesses, challenges and potential applications using this data. It covers also estimating population distribution, mobility patterns, types of activities in different parts of the city and analysing social networks formed through mobile networks. It also discusses techniques for analysing and processing CDR and depicts limitations of this type of data.
[31]	Analysis of GSM calls data for understanding user mobility behavior (2013)	This paper is about analysing user patterns (resident, commuter and visitor), which is possible because the authors had access to mobile phone data where the anonymous IDs did not change from day to day (as in our dataset), so they could analyse user movements, e.g., over a whole week.
[32]	Unveiling the complexity of human mobility by querying and mining massive trajectory data	This paper describes a very interesting outcome of an EU project called M-Atlas which has been developed to tackle questions regarding mobility patterns. It is using GPS data that have been collected from cars. The difference to Mobility Explorer is that (at the time of writing their paper) the application was not able to show a dynamic depiction of the changes over time but static maps. The application is an impressive data mining tool, called GeoPKDD.
[33]	Mobility, Data Mining and Privacy	This book covers everything concerning CDR and their visualisation. It is a collection of many papers and part of GeoPKDD project.
[34]	Mobility, Data Mining and Privacy: The GeoPKDD Paradigm	This paper develops a Mobility Manager. This work was part of EU FP6 GeoPKDD project and gives an overview of the project and research challenges. 17,000 vehicles with GPS trackers for one week were tracked in Milan, Italy. Data from GPS are selected for data mining and the authors mainly identify mobility patterns, O-D matrices and visualisation possibilities (mobility atlas). Mobility Explorer’s work uses passive CDR and provides more deep insights of datasets.
[35]	Development of origin–destination matrices using mobile phone call data	This work covers OD matrices by calculating cell tower-to-tower trips—O-D matrices for various time periods. CDR of 2.87 million users of Dhaka, Bangladesh over a period of one month combined with traffic counts at 13 different locations on 3 days of that month. There are 67 nodes and 215 links covering about 300 km ² with a population about 10.7 million. Only central part of Dhaka is studied. Mobile phone penetration rate in Dhaka is 90%. Studies 971.33 Million CDR.

Table A2. Cont.

Ref. Nr.	Title	Comparison to Mobility Explorer
[36]	Application of mobile phone location data in mapping of commuting patterns and functional regionalization: a pilot study of Estonia	This is also a highly relevant paper, though the data seem to be lower resolution than Mobility Explorer's.
[37]	* Mobile Phone Data to Describe Urban Practices: An Overview in the Literature	Authors cover various practices in literature and advocate on the benefits of mobile data in observing population distribution and mobility patterns in an urban landscape. But no internal working details about data, pre-processing, O-D matrices, interactive visualisation and examples are presented.
[38]	* Data from mobile phone operators: A tool for smarter cities?	This paper also covers a comprehensive review of CDR for spatio-temporal analysis (population distribution, mobility patterns, visual representation, social communities and network analysis in urban planning. This survey provides a high level point of view and is different from UrbanAPI where a real case study and detailed feasibility analysis of CDR is presented.
[39]	* Overview of the sources and challenges of mobile positioning data for statistics	The author covers details about mobile network infrastructure, active and passive data collection. CDR data examples are provided as well. Tourism and urban planning applications are discussed. Privacy issues and pre-processing issues are discussed, too. Data is visualised using various visual techniques. In contrast, UrbanAPI covers more features like specific O-D matrices and day-night population and interactive visualisation. Also, visualisation maps produced by the UrbanAPI application are providing more specific details of city districts which can be exported in raw data form for further analysis.
[40]	Discovering urban and country dynamics from mobile phone data with spatial correlation patterns (2014)	This paper is highly relevant but is focusing on static 2D map representations. The authors have been using a dataset from 2007 covering information on mobile phone movements over several days → the anonymous ID is not changing in this dataset on a daily basis.
[41]	Using Mobile Positioning Data to Model Locations Meaningful to Users of Mobile Phones	The data used in this paper is of significantly different nature as the data of UrbanAPI: In Estonia anonymous IDs seem not to change over the whole year, so it is possible to see each user's calling pattern over this period. The data in UrbanAPI had anonymous IDs changing every day (sic!) but since the data in UrbanAPI held all movements of users it was possible to conduct motion pattern analyses (the data of this specific paper covers only log entries when calls went out, so the authors could "only" count the calls and the position where the calls were made).

* These are survey papers and do not cover specific case study but provide a comprehensive review of existing work on CDR.

Table A3. Example of the O-D matrix extracted from the CDR data.

User_Count	CellID_Dest	Orig_CellID	Orig_Entity	Date	Orig_Timespan	Dest_Time
24	1kmN2816E4802	1kmN2816E4802	Raster	20120124	btw_0:0_4:0	24.01.2012 05:45
108	1kmN2818E4801	1kmN2818E4801	Raster	20120124	btw_0:0_4:0	24.01.2012 05:45
16	1kmN2819E4800	1kmN2818E4801	Raster	20120124	btw_0:0_4:0	24.01.2012 05:45
12	1kmN2817E4801	1kmN2818E4801	Raster	20120124	btw_0:0_4:0	24.01.2012 05:45
85	1kmN2814E4801	1kmN2814E4801	Raster	20120124	btw_0:0_4:0	24.01.2012 05:45
13	1kmN2813E4801	1kmN2814E4801	Raster	20120124	btw_0:0_4:0	24.01.2012 05:45
41	1kmN2814E4806	1kmN2814E4806	Raster	20120124	btw_0:0_4:0	24.01.2012 06:00
24	1kmN2816E4802	1kmN2816E4802	Raster	20120124	btw_0:0_4:0	24.01.2012 06:00
108	1kmN2818E4801	1kmN2818E4801	Raster	20120124	btw_0:0_4:0	24.01.2012 06:00
16	1kmN2819E4800	1kmN2818E4801	Raster	20120124	btw_0:0_4:0	24.01.2012 06:00
12	1kmN2817E4801	1kmN2818E4801	Raster	20120124	btw_0:0_4:0	24.01.2012 06:00
85	1kmN2814E4801	1kmN2814E4801	Raster	20120124	btw_0:0_4:0	24.01.2012 06:00
13	1kmN2813E4801	1kmN2814E4801	Raster	20120124	btw_0:0_4:0	24.01.2012 06:00
40	1kmN2814E4806	1kmN2814E4806	Raster	20120124	btw_0:0_4:0	24.01.2012 06:15
23	1kmN2816E4802	1kmN2816E4802	Raster	20120124	btw_0:0_4:0	24.01.2012 06:15
104	1kmN2818E4801	1kmN2818E4801	Raster	20120124	btw_0:0_4:0	24.01.2012 06:15
16	1kmN2819E4800	1kmN2818E4801	Raster	20120124	btw_0:0_4:0	24.01.2012 06:15
11	1kmN2817E4801	1kmN2818E4801	Raster	20120124	btw_0:0_4:0	24.01.2012 06:15
84	1kmN2814E4801	1kmN2814E4801	Raster	20120124	btw_0:0_4:0	24.01.2012 06:15
14	1kmN2813E4801	1kmN2814E4801	Raster	20120124	btw_0:0_4:0	24.01.2012 06:15
39	1kmN2814E4806	1kmN2814E4806	Raster	20120124	btw_0:0_4:0	24.01.2012 06:30
23	1kmN2816E4802	1kmN2816E4802	Raster	20120124	btw_0:0_4:0	24.01.2012 06:30
100	1kmN2818E4801	1kmN2818E4801	Raster	20120124	btw_0:0_4:0	24.01.2012 06:30
18	1kmN2819E4800	1kmN2818E4801	Raster	20120124	btw_0:0_4:0	24.01.2012 06:30
13	1kmN2817E4801	1kmN2818E4801	Raster	20120124	btw_0:0_4:0	24.01.2012 06:30
83	1kmN2814E4801	1kmN2814E4801	Raster	20120124	btw_0:0_4:0	24.01.2012 06:30
13	1kmN2813E4801	1kmN2814E4801	Raster	20120124	btw_0:0_4:0	24.01.2012 06:30
39	1kmN2814E4806	1kmN2814E4806	Raster	20120124	btw_0:0_4:0	24.01.2012 06:45
20	1kmN2816E4802	1kmN2816E4802	Raster	20120124	btw_0:0_4:0	24.01.2012 06:45
93	1kmN2818E4801	1kmN2818E4801	Raster	20120124	btw_0:0_4:0	24.01.2012 06:45
16	1kmN2819E4800	1kmN2818E4801	Raster	20120124	btw_0:0_4:0	24.01.2012 06:45
15	1kmN2817E4801	1kmN2818E4801	Raster	20120124	btw_0:0_4:0	24.01.2012 06:45
80	1kmN2814E4801	1kmN2814E4801	Raster	20120124	btw_0:0_4:0	24.01.2012 06:45
14	1kmN2813E4801	1kmN2814E4801	Raster	20120124	btw_0:0_4:0	24.01.2012 06:45

References

1. Khan, Z.; Ludlow, D.; Loibl, W.; Soomro, K. ICT enabled participatory urban planning and policy development: The UrbanAPI project. *Transform. Gov. People Process Policy* **2014**, *8*, 205–229. [[CrossRef](#)]
2. Khan, Z.; Liaquat, K.S.; Soomro, K. A framework for cloud-based context-aware information services for citizens in smart cities. *J. Cloud Comput.* **2014**, *3*, 1–17. [[CrossRef](#)]
3. Loibl, W.; Peters-Anders, J. Mobile phone data as source to discover spatial activity and motion patterns. In *GI_Forum 2012: Geovisualization, Society and Learning*; Von Jekel, T., Car, A., Strobl, J., Griesebner, G., Eds.; Wichmann Verlag: Berlin/Offenbach, Germany, 2012; pp. 524–533.
4. Khan, Z.; Anjum, A.; Soomro, K.; Muhammad, T. Towards cloud based big data analytics for smart future cities. *J. Cloud Comput.* **2015**, *4*, 2. [[CrossRef](#)]
5. Peters-Anders, J.; Loibl, W.; Züger, J.; Khan, Z.; Ludlow, D. Exploring Population Distribution and Motion Dynamics through Mobile Phone Device Data in Selected Cities-Lessons Learned from the UrbanAPI Project. In *Proceedings of the REAL CORP 2014 Proceedings/Tagungsband*, Vienna, Austria, 21–23 May 2014; pp. 871–876.
6. O'Neill, C.; Shutt, R. *Doing Data Science: Straight Talk from the Frontline*; O'Reilly Media, Inc.: Sebastopol, CA, USA, 2013.
7. Hevner, A.R.; March, S.T.; Park, J.; Ram, S. Design Science in Information Systems Research. *MIS Q.* **2004**, *28*, 75–105.
8. Hevner, A.R. A three cycle view of design science research. *Scand. J. Inf. Syst.* **2007**, *19*, 4.
9. Peffers, K.; Tuunanen, T.; Gengler, C.E.; Rossi, M.; Hui, W.; Virtanen, V.; Bragge, J. The Design Science Research Process: A Model for Producing and Presenting Information Systems Research. In *Proceedings of the First International Conference on Design Science Research in Information Systems and Technology (DESRIST 2006)*, Claremont, CA, USA, 24–25 February 2006; pp. 83–106.
10. March, S.T.; Storey, V.C. Design Science in Information Systems Discipline. *MIS Q.* **2008**, *32*, 725–730.
11. Michalopoulou, M.; Riihijärvi, J.; Mähönen, P. Studying the Relationships between Spatial Structures of Wireless Networks and Population Densities. In *Proceedings of the IEEE Conference on Global Telecommunications Conference*, Miami, FL, USA, 6–10 December 2010.
12. Readers, J.; Calabrese, F.; Sevtsuk, A.; Ratti, C. Cellular Census: Explorations in Urban Data Collection. *IEEE Pervasive Comput.* **2007**, *6*, 30–38. [[CrossRef](#)]
13. Vieira, M.R.; Frias-Martinez, V.; Oliver, N.; Frias-Martinez, E. Characterizing dense urban areas from mobile phone-call data: Discovery and social dynamics. In *Proceedings of the 2nd IEEE Conference on Social Computing (SocilCom 2010)*, Minneapolis, MN, USA, 20–22 August 2010; pp. 241–248.
14. Isaacman, S.; Becker, R.; Caceres, R.; Kobourov, S. Identifying Important Places in People's Lives from Cellular Network Data. In *Proceedings of the 9th International Conference on Pervasive Computing (Pervasive)*, San Francisco, CA, USA, 12–15 June 2011; pp. 133–151.
15. Phithakitnukoon, S.; Horanont, T.; Di Lorenzo, G.; Shibasaki, R.; Ratti, C. Activity-aware map: Identifying human daily activity pattern using mobile phone data. In *Proceedings of the Human Behavior Understanding*, Istanbul, Turkey, 22 August 2010; pp. 14–25.
16. Palmer, J.R.; Espenshade, T.J.; Bartumeus, F.; Chung, C.Y.; Ozgencil, N.E.; Li, K. New approaches to human mobility: Using mobile phones for demographic research. *Demography* **2013**, *503*, 1105–1128. [[CrossRef](#)] [[PubMed](#)]
17. Caceres, N.; Wideberg, J.P.; Benitez, F.G. Deriving origin–destination data from a mobile phone network. *IET Intell. Transp. Syst.* **2007**, *1*, 15–26. [[CrossRef](#)]
18. Steenbruggen, J.; Borzacchiello, M.T.; Nijkamp, P.; Scholten, H. Mobile phone data from GSM networks for traffic parameter and urban spatial pattern assessment: A review of applications and opportunities. *Geo J.* **2013**, *78*, 223–243. [[CrossRef](#)]
19. Blondel, V.D.; Decuyper, A.; Krings, G. A survey of results on mobile phone datasets analysis. *EPJ Data J.* **2015**, *4*, 10. [[CrossRef](#)]
20. Nanni, M.; Trasarti, R.; Furletti, B.; Gabrielli, L.; Van Der Mede, P.; De Bruijn, J.; De Romph, E.; Bruil, G. Transportation Planning Based on GSM Traces: A Case Study on Ivory Coast. In *Citizen in Sensor Networks*; Nin, J., Villatoro, D., Eds.; Springer: Cham (ZG), Switzerland, 2014; pp. 15–25.

21. Raslan, H.; Elragal, A. Inferring Urban Population Distribution from GSM Data: An Experimental Case Study. In Proceedings of the GEOProcessing 2015: The Seventh International Conference on Advanced Geographic Information Systems, Applications, and Services, Lisbon, Portugal, 22–27 February 2015.
22. Dash, M.; Nguyen, H.L.; Hong, C.; Yap, G.E.; Nguyen, M.N.; Li, X.; Krishnaswamy, S.P.; Decraene, J.; Antonatos, S.; Wang, Y.; et al. Home and Work Place Prediction for Urban Planning Using Mobile Network Data. In Proceedings of the 15th International Conference on Mobile Data Management, Brisbane, Australia, 14–18 July 2014; Volume 2, pp. 37–42.
23. Dash, M.; Chua, G.G.; Nguyen, H.-L.; Yap, G.-E.; Hong, C.; Li, X.; Krishnaswamy, S.P.; Decraene, J.; Shi-Nash, A. An Interactive Analytics Tool for Understanding Location Semantics and Mobility of Users Using Mobile Network Data. In Proceedings of the 15th International Conference on Mobile Data Management, Brisbane, Australia, 14–18 July 2014; Volume 1, pp. 345–348.
24. Wesolowski, A.; Eagle, N.; Noor, A.M.; Snow, R.W.; Buckee, C.O. The impact of biases in mobile phone ownership on estimates of human mobility. *J. R. Soc. Interface* **2013**, *10*. [[CrossRef](#)] [[PubMed](#)]
25. Frias-Martinez, V.; Virseda, J. On the relationship between socio-economic factors and cell phone usage. In Proceedings of the Fifth International Conference on Information and Communication Technologies and Development, Atlanta, GA, USA, 12–15 March 2012; pp. 76–84.
26. Isaacman, S.; Becker, R.; Cáceres, R.; Martonosi, M.; Rowland, J.; Varshavsky, A.; Willinger, W. Human Mobility Modeling at Metropolitan Scales. In Proceedings of the 10th International Conference on Mobile Systems, Applications, and Services, Low Wood Bay, Lake District, UK, 25–29 June 2012; pp. 239–252.
27. Khan, Z.; Ludlow, D.; Cáceres, S. Evaluating a collaborative IT based research and development project. *Eval. Progr. Plan.* **2013**, *40*, 27–41. [[CrossRef](#)] [[PubMed](#)]
28. IMSI (International Mobile Subscriber Identity)—Assignment and Management Guidelines and Procedures. V.12.0; Telcordia Technologies, Inc.: Piscataway, NJ, USA, 2010.
29. Andrienko, G.; Andrienko, N.; Demsar, U.; Dransch, D.; Dykes, J.; Fabrikant, S.I.; Jern, M.; Kraak, M.J.; Schumann, H.; Tominski, C. Space, time and visual analytics. *Int. J. Geogr. Inf. Sci.* **2010**, *24*, 1577–1600. [[CrossRef](#)]
30. Clabrese, F.; Ferrari, L.; Blondel, V. Urban Sensing using Mobile Phone Network Data: A survey of research. *ACM Comput. Surv.* **2015**, *47*, 25. [[CrossRef](#)]
31. Furletti, B.; Gabrielli, L.; Renso, C.; Rinzivillo, S. Analysis of GSM calls data for understanding user mobility behavior. In Proceedings of the 2013 IEEE International Conference on Big Data, Silicon Valley, CA, USA, 6–9 October 2013.
32. Giannotti, F.; Nanni, M.; Pedreschi, D.; Pinelli, F.; Renso, C.; Rinzivillo, S.; Trasarti, R. Unveiling the complexity of human mobility by querying and mining massive trajectory data. *VLDB J.* **2011**, *20*, 695. [[CrossRef](#)]
33. Giannotti, F.; Pedreschi, D. (Eds.) *Mobility, Data Mining and Privacy: Geographic Knowledge Discovery*; Springer: Berlin/Heidelberg, Germany, 2007.
34. Giannotti, F.; Trasarti, R. Mobility, Data Mining and Privacy: The GeoPKDD Paradigm. In Proceedings of the 14th SIAM Conference on Mathematics for Industry (MI'09), San Francisco, CA, USA, 9–11 October 2009; pp. 10–18.
35. Iqbal, M.S.; Choudhury, C.F.; Wang, P.; Gonzalez, M.C. Development of origin-destination matrices using mobile phone call data. *Transp. Res. Part C* **2014**, *40*, 63–74. [[CrossRef](#)]
36. Novak, J.; Ahas, R.; Aasa, A.; Silm, S. Application of mobile phone location data in mapping of commuting patterns and functional regionalization: A pilot study of Estonia. *J. Maps* **2013**, *9*, 10–15. [[CrossRef](#)]
37. Pucci, P.; Manfredini, F.; Tagliolato, P. Mobile Phone Data to Describe Urban Practices: An Overview in the Literature. In *Mapping Urban Practices through Mobile Phone Data*; Springer: Cham (ZG), Switzerland, 2015; pp. 13–25.
38. Steenbruggen, J.; Tranos, E.; Nijkamp, P. Data from mobile phone operators: A tool for smarter cities? *Telecommun. Policy* **2015**, *39*, 335–346. [[CrossRef](#)]
39. Tiru, M. Overview of the sources and challenges of mobile positioning data for statistics. In Proceedings of the International Conference on Big Data for Official Statistics, Beijing, China, 28–30 October 2014.

40. Trasart, R.; Olteanu-Raimond, A.M.; Nanni, M.; Couronne, T.; Furletti, B.; Giannotti, F.; Smoreda, Z.; Ziemlicki, C. Discovering urban and country dynamics from mobile phone data with spatial correlation patterns. *Telecommun. Policy* **2015**, *39*, 347–362. [[CrossRef](#)]
41. Ahas, R.; Silm, S.; Jarv, O.; Saluveer, E.; Tiru, M. Using Mobile Positioning Data to Model Locations Meaningful to Users of Mobile Phones. *J. Urban Technol.* **2010**, *17*, 3–27. [[CrossRef](#)]



© 2017 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).