
Peters-Anders\textsuperscript{1,*}, J., Khan, Z.\textsuperscript{2}, Loibl, W\textsuperscript{3}, Augustin, H\textsuperscript{3}, Breinbauer, A\textsuperscript{1}.

\textsuperscript{1} Austrian Institute of Technology, Energy Department, Sustainable Buildings and Cities Unit, Vienna, Austria. Tel: +43 664 815 7995, Email: Jan.Peters-Anders@ait.ac.at, Wolfgang.Loibl@ait.ac.at, arno.breinbauer@edu.fh-kaernten.ac.at

\textsuperscript{2} University of the West of England, Bristol, UK, Tel: +44 1173287216, Email: Zaheer2.Khan@uwe.ac.uk

\textsuperscript{3} Municipal Department for Urban Development and Planning (MA 18), City of Vienna, Tel: +43 (1) 4000-88714, Email: helmut.augustin@wien.gv.at

\textsuperscript{*} Correspondence: Jan.Peters-Anders@ait.ac.at

Academic Editor: name

Received: date; Accepted: date; Published: date

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, 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 which 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 validation and critically reflects on the evaluation results of the citizens’ motion exploration which reveal great potential of mobile phone data in smart city planning but also depicts 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 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) which are independent of the 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 of urban population city planners are always looking for evidence based information sources to effectively and efficiently plan and utilise public resources. And, 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 built environment and public transport to improve the quality of life of citizens.

However, population distribution is since long detected through census activities. These data are important but provide insufficient information on 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 daily socio-economic activities of the citizens which require them to travel from one place to another. As census data give only a coarse image of the population distribution and its changes during the day until now, mobile phone location data (i.e. CDR) turns out to be a unique and appreciated data 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 time, location and further activity information for each mobile device a.k.a 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, update cell information). These data can be used to detect patterns of population distribution and
perform mobility analyses. However, the methods and means 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 there is a mistrust applying these data which 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 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 in mobile cells equipped with cell towers (bearing antennas). The technical infrastructure observes continuously 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 require 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 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) which 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 effect the process of determining mobility patterns. All these constraints require a thorough investigation by 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 Loibl, W., and Peters-Anders, J. (2012, p527) [3] highlighted the issues related to acquisition of mobile data and its potential benefits in city planning. This work was complemented in Peters-Anders, J., et al. (2014, p6) [5] with basic experiments as proof of concept by using small sample of mobile data to assess its feasibility in 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 (Table 3a). 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) which affect fulfilling new requirements of city planners. 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 (Sept 2011- Dec 2014) aiming at supporting urban
planning through ICT tools. In UrbanAPI, the 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 potential benefits and limitations of
such data to support city planning; ii) the obtained raw data of Vienna allows investigating
the range of possibilities of the data and its potential for city planning and its 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 shared based on experiences of
Vitoria-Gasteiz case in section 8 to reflect on the suitability of Mobility Explorer for smaller
sized cities.

1.3 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 Design Science Research Methodology (DSRM) by Hevner et
al. [25][29] which is a well-accepted method in information systems research (e.g Peffers et
al. (2006) or March and Storey (2008)) [26][27]. 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 (Table 3a below) was obtained from the
city administration and integrated with mobile phone data to generate 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.
1.4 Paper structure

In the remainder of this paper, 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 to 150%, e.g. in Austria 8 million citizens hold around 12 million mobile device subscriptions, resulting in a subscriber/population ratio of 150%¹. Michalopoulou et al. (2010) [7] have proven that there exists 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 literature, there exist some examples which provide good insight about the use of mobile data to derive population distribution and mobility patterns that can be used in various urban applications ([8], [9], [10], [11], [12], [13]). Steenbruggen et al., (2013) [14] work can be considered as 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’s et al. (2013) [14] article provide basic research questions which actually can 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. (2015) [15] where authors question the suitability of the mobile data e.g. they argue that datasets are noisy, some links appear there by chance, while other 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 during the pre-processing of data is needed to be able to use them within Mobility Explorer.

Similarly, Nanni M et al. (2014) [16] present a case study (Ivory Coast) where 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 systematic mobility behaviour of people between

frequently visited locations for areas that lack information on mobility. Raslan and Elragal (2015) 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 dynamics of population distributions and mobility patterns using a visual interactive tool and deduce possible reasoning. Dash et al. (2014a) [18] worked on new algorithms to predict 25% more accurate locations (i.e. home and workplaces during weekdays and weekends) using GSM data when phones are inactive. Experiments carried out in Dash et al. (2014b) [19] and cross-validation of results from GSM mobile data and open statistical data provides promising results to extract population distribution patterns through an interactive application. In both 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 the 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 Table 4 in the Appendix 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 which should be considered when developing a smart Mobility Explorer application for cities. Figure 1 depicts five main elements of the proposed framework.
The first element, Stakeholders & Requirements emphasises the need of 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 data sets required to develop appropriate solutions. The second element, Data & Governance covers what data sets are available through different departments of cities? What auxiliary data is available from third parties? In which 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 is this data accessible etc.? These data sets 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 data set is sufficient to fulfil the city needs the third element, Processes & Algorithms defines certain alignment processes to map available data sets 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 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 which 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 above elements of the proposed framework with suitable examples.

In order to operationalize 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 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 get insights about a city’s socio-economic dynamics which -with conventional methods- is either difficult or very expensive to get at neighbourhood, city or city-region scale. These new insights provide evidence based information for better planning and decision making. There are various conventional methods which 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:
Table 1: Conventional data collection methods for population distribution and mobility patterns

<table>
<thead>
<tr>
<th>Data collection Method</th>
<th>Benefits</th>
<th>Limitations</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Population Census and Register Data:</strong> This is the most common method applied across the world. Population census or surveys are used to collect this data</td>
<td>Census data is highly representative. It is not a sample but a survey covering the entire population</td>
<td>In different countries collected content vary. Also, migration movements and commuting to work destinations are collected through static surveys. But census surveys are now often replaced by register counts</td>
</tr>
<tr>
<td><strong>Counting and Sensor Data:</strong> This static data collection method uses traffic counters or special sensors to measure traffic volume on specific roads</td>
<td>It is the best estimated information about traffic volume on a specific point on the traffic network</td>
<td>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 in underground lines provides at least vague information about origins and destinations</td>
</tr>
<tr>
<td><strong>Polling:</strong> This is another type of survey that is used to collect mobility data</td>
<td>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</td>
<td>It is never representative at high spatial resolution due to limited sample sizes</td>
</tr>
</tbody>
</table>

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 in 2015 around 5.5 million subscribers, resulting in a market share of around 42%. The abovementioned numbers prove as satisfactory representativeness of the dataset to describe the entire population distribution dynamics, letting assume 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. (2013) [20], Frias-Martinez (2012) [21] and Blondel et al. (2015) [15] have advocated that differential mobile ownership biases do not seem to have much effect on mobility patterns within particular regions. This suggests that social bias of the clients of certain providers is appearing to be 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 which ensure sufficient representativeness for mobility analysis based on mobile phones.

From a technical perspective the location detection of a cell phone is done using different methods e.g. triangulation and cell tower aggregation. Since movement of customers causes 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 which 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 meters in city centres to several kilometres in rural areas. 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, signal oscillation between cell towers may indicate pseudo movements of users which 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 finally 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 (little) disadvantage but this can be neglected if the cells are small which is mostly the case in dense urban environments.

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

---

3 Example of cell tower maps in Austria is accessible from [www.senderkataster.at](http://www.senderkataster.at) Last accessed: 04 Apr 2016
Table 2: Mobile phone data collection methods for population distribution and mobility

<table>
<thead>
<tr>
<th>Data collection Method</th>
<th>Benefits</th>
<th>Limitations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cell or Cell Triangulation Data</td>
<td>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.</td>
<td>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.</td>
</tr>
<tr>
<td>Global Positioning System (GPS) Data</td>
<td>It provides high spatial and temporal resolution.</td>
<td>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, 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.</td>
</tr>
<tr>
<td>Cell Data Records (CDR): This method is used to collect data based on cell phone usage or activities.</td>
<td>This data collection method works passively and end users do not need to install any new app that results in efficient battery consumption. Privacy concerns can be handled by cell service providers by anonymisation techniques. Also, data is highly representative at high spatial and temporal resolution.</td>
<td>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.</td>
</tr>
</tbody>
</table>
The amount of data collected by mobile service providers is enormous and for large telecom operators it can be in few 100s of GBytes to TBytes on a weekly basis. 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.

**Table 3a**: Data sets used within the Mobility Explorer (ME) application

<table>
<thead>
<tr>
<th>Usage of Dataset in Mobility Explorer</th>
<th>Dataset Type</th>
<th>Data Type</th>
<th>Data Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Basemaps for orientation purposes</td>
<td>Quarter level</td>
<td>Vector</td>
<td>Planning authorities</td>
</tr>
<tr>
<td>Basemaps for orientation purposes</td>
<td>City level</td>
<td>Raster</td>
<td>Planning authorities</td>
</tr>
<tr>
<td>Background layer for selecting analysis region</td>
<td>Census units (at least 1 per km² raster cell)</td>
<td>Tables Raw</td>
<td>Planning authorities National statistics</td>
</tr>
</tbody>
</table>

4 This also depends on size of the country and market share of the mobile operator. In Austria it can be 100s of GB per week but in Germany it may be 10 times more.
<table>
<thead>
<tr>
<th>Urban Topography</th>
<th>Basemap for orientation and visual analysis purposes</th>
<th>Satellite / ortho-images</th>
<th>X</th>
<th>Planning authorities</th>
</tr>
</thead>
<tbody>
<tr>
<td>Population (Census)</td>
<td>Basis for statistical projection of mobile phone users to total population</td>
<td>Population Total</td>
<td>X</td>
<td>National statistics</td>
</tr>
<tr>
<td>Mobile Phone Data</td>
<td>Basis for visual analysis and retrieving mobility analysis data</td>
<td>Raw data stream</td>
<td>X</td>
<td>Service providers</td>
</tr>
<tr>
<td></td>
<td>Basis for visual analysis within ME</td>
<td>Preprocessed data</td>
<td>X</td>
<td>Own calculations</td>
</tr>
<tr>
<td></td>
<td>Basemap for orientation and visual analysis purposes (only available in Vitoria-Gasteiz application)</td>
<td>Cellular antenna locations</td>
<td>X</td>
<td>Service providers,</td>
</tr>
<tr>
<td></td>
<td>Basis for statistical projection of mobile phone users to total population</td>
<td>Mobile Phone Total</td>
<td>X</td>
<td>National statistics</td>
</tr>
</tbody>
</table>

Table 3b: Data sets to be considered for future applications

<table>
<thead>
<tr>
<th>Usage to consider in the future</th>
<th>Dataset Type</th>
<th>Data Type</th>
<th>Data Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Basemaps</td>
<td>Basemap for orientation purposes</td>
<td>Block to block clusters</td>
<td>Vector</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Raster</td>
</tr>
<tr>
<td>Urban Topography</td>
<td>Basemap for orientation purposes</td>
<td>Transportation network (road network)</td>
<td>Vector</td>
</tr>
<tr>
<td></td>
<td>Basemap for orientation and visual analysis purposes</td>
<td>Public transport network (railways, light rails, trams, sub-ways, bus routes, stations)</td>
<td>Vector</td>
</tr>
</tbody>
</table>
### Table: Mobile Phone Data Uses and Sources

<table>
<thead>
<tr>
<th>Source / Data Type</th>
<th>Description</th>
<th>Available for bias detection</th>
<th>Complexity</th>
<th>Correlation</th>
<th>Source Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Population Data (Census)</td>
<td>Mobile phones by age classes (0-15yrs, 15-45yrs, 45-65yrs, 65+yrs)</td>
<td>If available useful</td>
<td>X</td>
<td>X</td>
<td>National statistics</td>
</tr>
<tr>
<td></td>
<td>Population by gender, by social class characteristics by education</td>
<td></td>
<td></td>
<td></td>
<td>National statistics</td>
</tr>
<tr>
<td></td>
<td>Temporary population (students, tourists etc)</td>
<td></td>
<td></td>
<td>X</td>
<td>National statistics</td>
</tr>
<tr>
<td></td>
<td>Basemap for orientation and visual analysis purposes.</td>
<td>X</td>
<td></td>
<td></td>
<td>Local authorities</td>
</tr>
<tr>
<td>Households</td>
<td>Basis for statistical projection of mobile phone.</td>
<td></td>
<td>X</td>
<td></td>
<td>National statistics</td>
</tr>
<tr>
<td>Workplaces</td>
<td>Basis for statistical projection of mobile phone users.</td>
<td></td>
<td>X</td>
<td></td>
<td>National statistics</td>
</tr>
</tbody>
</table>

Since the CDR data is 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. by map-reduce using Hadoop to divide larger files into smaller size files and process them on Hadoop worker nodes. However, then synthesizing results for 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 (cp. e.g. [28]) 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 o’clock 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 o’clock p.m. any more. For this reason Mobility
Explorer is focusing on dynamic visualisations of movements of bigger groups of mobile phone users like e.g. commuters to let e.g. city administrations check for newly arising streams of commuters that have developed only recently and which have not been detectable in the past since the tools to detect this have been missing.

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. But the existing data on 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 as an 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 high interest to investigate the potential of using mobile phone data to visualise and analyse diurnal population distribution and public motion dynamics that can potentially be used to achieve the above goals.

Vienna considers mobile phone data as an entirely new source of information concerning socio-economic 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 a better insight in 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 is appreciated as an extension to the 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 at which daytime in which district and/or which 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 o’clock. At the weekend for 10:00 and 15:00 o’clock.

Vienna is also interested in a visual interactive tool where 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 o’clock) 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 space. The spatial resolution of the provided data are 1,000 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 server side consists of a PostgreSQL/PostGIS database. Mobility Explorer features and
a Tomcat servlet as well as a Geoserver instance to visualise the pre-processed motion data derived
from the mobile phone logs. These tools are commonly used in geo-spatial application development.
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, base maps 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 was acquired from a major mobile communication service
provider. This allowed 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 data set from the raw CDR data is shown in Figure 4.
Figure 4: Processing Steps: From Raw CDR Data to the Dynamic Visualisation Data Set

7. Motion Exploration: Vienna Application outcomes – Outcomes & Evaluation Element

a. 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 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 cell towers in the West of Austria reported false positions. A similar issue was detected in the Vitoria-Gasteiz data set due to its low density of cell antennas (see Figure 14). Such errors – if not detected - can lead city planners to make false assumptions of 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 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. For the Vienna application, a 15 minutes time step was selected to explore and view temporal dynamics. This 15 minutes time step was acceptable by 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 chain of trip positions of unique devices.

![Figure 5](Image)

**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.

However, during the chaining process further position errors were detected i.e. mobile devices where observed as if they were “jumping” to far distance locations within the 15 minute time steps. It turned out that these pseudo movements occur due to load balancing reasons i.e. heavily
used cell towers redirect devices which exceed the cell tower’s carrying capacity to other cell
towers with less data traffic. This made it sometimes 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 which would require impossible travel speed to reach the cell within
this time where 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 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).

b. 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 has been coming from one single provider there was also the need for
checking the feasibility to project the mobile phone user data to the whole city population.
This has been done by comparing the 1km² grid cell inhabitants of the national statistical
institute with the aggregated 1km² grid mobile phone user numbers for the time period of
0-6 o’clock (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,
like e.g. policemen, taxi drivers, clinic staff etc.)). The map in Figure 6 shows the result:
The statistical number of inhabitants per 1km² cell (depicted as coloured “dots” in the
map) matches the spatial distribution of the mobile devices (depicted as coloured
“rectangles”): Where there is 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 maps results in a statistical graph showing the relation of
statistical 1km² population to 1km² “sleeping population” from the mobile phone user data
set: The value of $R^2 = 0.5533$ shows a quite good correlation between the observed mobile
devices and the statistical population in the 1km² grid cells which leads to the assumption
that it is indeed feasible to calculate a coefficient for each source cell:

$$\text{coefficient}_{\text{source cell } x} = \frac{\text{statistical inhabitants}_{\text{source cell } x}}{\text{sleeping population}_{\text{source cell } x}}$$
Figure 6: Visual comparison between 1km² statistical population of Vienna (dot signature) vs.
1km² mobile phone user densities between 0-6 o’clock (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 with this coefficient to calculate the
representing statistical population coming from the source cell.

Figure 7: Statistical comparison between 1km² statistical population of Vienna vs. 1km² mobile
phone user densities between 0-6 o’clock (R²=0.5533)

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

- **Dynamic mapping of diurnal motion dynamics for interactively selected cells (origins):** This
functionality lets the user choose any “source cell” (the red rectangle in the map) which 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 the 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](image)

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

![Figure 9](image)

**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 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 hours). The left hand side figure depicts 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 side figure.
The above interactive and dynamic population distribution and mobility analysis at city scale can also be saved as origin destination matrices (an example in Appendix) which provide new information 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 analyses. This information is also useful to analyse population dynamics at 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 allows 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 includes all trips (work, leisure, shopping, education, etc.) while Austrian census data includes trips to work only. Looking at the high percentages depicted in the legend we recognize how remarkably interwoven the Vienna Region is. So this comprehensive dataset can substantially contribute to identifying the functional urban region – an often discussed research question among regional planners. Another aspect is the distinct impact of high level transport infrastructure on mobility behaviour of people. This applies for motorways (black lines in the map) as well as for 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 huge impact on achieving Green House Gas (GHG) emission targets and improving quality of life of 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 (Figure 11).
2017, 8, x FOR PEER REVIEW 22 of 33

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)

12 and Figure 13). Figure 12 depicts that people living in the inner districts cause less kilometres travelled and this indicates that majority of people are not taking long journeys for their routine activities.
In contrast to Figure 12, Figure 13 depicts that people who live in suburban areas have to travel to inner districts for routine activities. For instance, it depicts that people from Southern suburbs have high percentage of travel as compared to suburbs in North, East or West of City of Vienna. Such information at such a large scale provides stimulus to initiate new planning projects including creating job opportunities near 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)

c. 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 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 to assess 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 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 benefits and functionality of motion exploration and due to space limitations here we briefly present usability evaluation results.

Most of the evaluators found the application easy to understand and use but also indicated that resolution of maps can 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 running simulations showing dynamically changing
space occupation for various zones of the selected city. On overall performance, 80% evaluators agree that response time to user queries was reasonable but 20% disagree. This is mainly due to 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 for this application so that response time to multiple users’ queries is acceptable. All evaluators disagree that system is complex. This may be the reason where 40% to 60% evaluators agree that they would like to use this system frequently but 40% responses are neutral which may be related to specific needs of this system in their organisations’ core businesses. Up to 60% evaluators agree and 40% are 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 specific operation is performed. For example while the simulation is running to show the population distribution at different hours of a day, other features are still accessible and can be activated to interrupt the simulation. Up to 80% disagree that system is cumbersome to use and 20% neutral response may be attributed to need for intuitiveness and easy to use functions using application GUI. About 80% evaluators disagree that they had to learn too many things before they could get going with the system but only 20% agreed. These results suggest that 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 skillset of end users and there is potential to improve the application GUI for further intuitiveness and usability. Overall evaluators found Mobility Explorer easy to use and produce required outcomes (e.g. O-D matrices, high-resolution images, etc.) that can be further utilised in urban and transport planning processes. Likewise to usability, positive responses were obtained for functionality and benefits related evaluation criteria.

8. Lessons learned and recommendations

As presented through the previous sections, the Mobility Explorer framework provides all necessary elements which are needed to develop a visual tool using data from variety of sources that can satisfy real needs of city stakeholders. In this process of applying the framework on 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 before the CDR data had been visualised and thoroughly checked that certain questions the participating cities 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) enough 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 lot of inaccurate location information for mobile devices and hence proved to be less usable for motion exploration.
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 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.

The level of details 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.

Mobile data privacy is a critical issue. All companies who provide CDR data (for example, in case of Mobility Explorer from Austria, Italy and Spain) have strict policies about sharing such data e.g. some provide aggregated numbers which do not require attention for 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 is fully anonymised. Many companies do not offer data provision service and the data is 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 daily basis hinders long term observations of a single entity to protect privacy. Further details can be found in IMSI [22]. 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 taking when working with these datasets by complying to 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 Direction

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

- Visual interactive maps which 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. heat maps, accuracy of CDR maps etc.

In addition to 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 such data at city and 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 high or low cell density or technical infrastructure of mobile service providers play a crucial role in determining effectiveness of CDR in visualising diurnal population distribution and mobility dynamics.

The Mobility Explorer evaluation results are very positive in a sense that they indicate strengths and identify potential future directions of development. The functional appropriateness is validated as well as 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 city (or 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 details in the available CDR which restricts fulfilling 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 gaining insights of 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 question would be to combine mobile phone data
with social network information to increase the reliability and information density of the resulting
combined data sets [15].

**Author Contributions**

Jan Peters-Anders and Arno Breinbauer conducted the pre-processing and analysis of the raw mobile
phone data and programmed the data base routines to do so. Jan Peters-Anders furthermore
programmed the web application which 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.

Helmut Augustin was responsible for testing the results of the analysis within the city administration
of Vienna.

**Acknowledgement**

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

**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**

Table 4: Matrix of references dealing with the topic CDR data analysis and visualisation:

<table>
<thead>
<tr>
<th>Ref. Nr.</th>
<th>O-D Matrix</th>
<th>Temporal and Spatial Population Distribution</th>
<th>Temporal and Spatial Mobility Patterns</th>
<th>Data Issues</th>
<th>Information overlays</th>
<th>Pre-processing</th>
<th>Interactive Maps</th>
<th>2D Maps</th>
<th>Privacy</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>District-to-Point</td>
<td>Weekdays (or Weekends)</td>
<td>15 minutes</td>
<td>Cell-to-Cell (or City-Region)</td>
<td>Mode detection</td>
<td>census</td>
<td>Demographics</td>
<td>Gender analysis</td>
</tr>
<tr>
<td>[28]</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>[29]</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>[30]</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>[31]</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>[32]</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>[33]</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>[34]</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>[35]</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
</tbody>
</table>
Table 5: Summary of related work in comparison to Mobility Explorer (cp. also Table 4 above)

<table>
<thead>
<tr>
<th>Ref. Nr.</th>
<th>Title</th>
<th>Comparison to Mobility Explorer</th>
</tr>
</thead>
<tbody>
<tr>
<td>[28]</td>
<td>Human Mobility Modeling at Metropolitan Scales (2012)</td>
<td>This paper discusses a procedure called WHERE (&quot;Work and Home Extracted REgions&quot;) It describes a model to calculate probability distributions.</td>
</tr>
<tr>
<td>[29]</td>
<td>Using Mobile Positioning Data to Model Locations Meaningful to Users of Mobile Phones</td>
<td>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(!) but since the data in UrbanAPI held all movements of users it was possible to conduct motion pattern analyses (the data of this paper had only log entries when calls went out, so they could &quot;only&quot; count the calls and the position where the calls were made).</td>
</tr>
<tr>
<td>[30]</td>
<td>Space, time and visual analytics (2010)</td>
<td>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)</td>
</tr>
<tr>
<td>[31]</td>
<td>* A survey of results on mobile phone datasets analysis</td>
<td>Provides a survey of CDR used for exploratory social network analysis or mobile communities formed based on call records. This helped 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. 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.</td>
</tr>
<tr>
<td>Reference</td>
<td>Title</td>
<td>Description</td>
</tr>
<tr>
<td>-----------</td>
<td>------------------------------------------------------------------------</td>
<td>---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>[32]</td>
<td>Urban Sensing Using Mobile Phone Network Data: A Survey of Research</td>
<td>A highly relevant survey paper. It outlines CDR and discusses strengths and weaknesses, challenges and potential applications using this data. Talks about 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.</td>
</tr>
<tr>
<td>[33]</td>
<td>Analysis of GSM calls data for understanding user mobility behavior (2013)</td>
<td>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 over a whole week e.g.</td>
</tr>
<tr>
<td>[34]</td>
<td>Unveiling the complexity of human mobility by querying and mining massive trajectory data</td>
<td>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 has 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. --&gt; GeoPKDD</td>
</tr>
<tr>
<td>[35]</td>
<td>Mobility, Data Mining and Privacy</td>
<td>This book covers everything concerning CDR and their visualisation. It is a collection of many papers and part of GeoPKDD project.</td>
</tr>
<tr>
<td>[36]</td>
<td>Mobility, Data Mining and Privacy: The GeoPKDD Paradigm</td>
<td>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 form GPS is selected for data mining - mainly identifies mobility patterns, O-D matrices, visualisation (mobility atlas). Mobility Explorer’s work uses passive CDR and provides more deep insights of datasets.</td>
</tr>
<tr>
<td>[37]</td>
<td>Development of origin–destination matrices using mobile phone call data</td>
<td>This work covers OD matrices by calculating cell tower-to-tower trips - OD 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 300km² 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.</td>
</tr>
<tr>
<td>[38]</td>
<td>Application of mobile phone location data in mapping of commuting patterns and functional</td>
<td>This is also a highly relevant paper. The data seems to be of lower resolution than Mobility Explorer’s</td>
</tr>
<tr>
<td>Citation</td>
<td>Title</td>
<td>Summary</td>
</tr>
<tr>
<td>----------</td>
<td>----------------------------------------------------------------------</td>
<td>---------</td>
</tr>
<tr>
<td>[39]</td>
<td>Mobile Phone Data to Describe Urban Practices: An Overview in the Literature</td>
<td>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.</td>
</tr>
<tr>
<td>[40]</td>
<td>Data from mobile phone operators: A tool for smarter cities?</td>
<td>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.</td>
</tr>
<tr>
<td>[41]</td>
<td>Overview of the sources and challenges of mobile positioning data for statistics</td>
<td>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. Also, 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.</td>
</tr>
<tr>
<td>[42]</td>
<td>Discovering urban and country dynamics from mobile phone data with spatial correlation patterns (2014)</td>
<td>This paper is highly relevant but is focusing on static 2D map representations. The have been using a dataset from 2007 covering information on mobile phone movements over several days --&gt; anonymous ID is not changing in this data set on a daily basis.</td>
</tr>
</tbody>
</table>

* these are survey papers and do not cover specific case study but provide a comprehensive review of existing work on CDR
Table 6: Example of the O-D matrix extracted from the CDR data:

<table>
<thead>
<tr>
<th>user_count</th>
<th>cellid_dst</th>
<th>origin_cellid</th>
<th>o_entry_date</th>
<th>orig_timespan</th>
<th>dest_time</th>
</tr>
</thead>
<tbody>
<tr>
<td>13</td>
<td>10m28134f801</td>
<td>10m2814f4801</td>
<td>Raster</td>
<td>20120124</td>
<td>bbw_00_0</td>
</tr>
<tr>
<td>4</td>
<td>10m28134f801</td>
<td>10m2814f4801</td>
<td>Raster</td>
<td>20120124</td>
<td>bbw_00_0</td>
</tr>
<tr>
<td>4</td>
<td>10m28134f801</td>
<td>10m2814f4801</td>
<td>Raster</td>
<td>20120124</td>
<td>bbw_00_0</td>
</tr>
<tr>
<td>3</td>
<td>10m28134f801</td>
<td>10m2814f4801</td>
<td>Raster</td>
<td>20120124</td>
<td>bbw_00_0</td>
</tr>
<tr>
<td>1</td>
<td>10m28134f801</td>
<td>10m2814f4801</td>
<td>Raster</td>
<td>20120124</td>
<td>bbw_00_0</td>
</tr>
<tr>
<td>1</td>
<td>10m28134f801</td>
<td>10m2814f4801</td>
<td>Raster</td>
<td>20120124</td>
<td>bbw_00_0</td>
</tr>
<tr>
<td>41</td>
<td>10m28134f801</td>
<td>10m2814f4801</td>
<td>Raster</td>
<td>20120124</td>
<td>bbw_00_0</td>
</tr>
<tr>
<td>2</td>
<td>10m28134f801</td>
<td>10m2814f4801</td>
<td>Raster</td>
<td>20120124</td>
<td>bbw_00_0</td>
</tr>
<tr>
<td>1</td>
<td>10m28134f801</td>
<td>10m2814f4801</td>
<td>Raster</td>
<td>20120124</td>
<td>bbw_00_0</td>
</tr>
<tr>
<td>1</td>
<td>10m28134f801</td>
<td>10m2814f4801</td>
<td>Raster</td>
<td>20120124</td>
<td>bbw_00_0</td>
</tr>
<tr>
<td>1</td>
<td>10m28134f801</td>
<td>10m2814f4801</td>
<td>Raster</td>
<td>20120124</td>
<td>bbw_00_0</td>
</tr>
<tr>
<td>6</td>
<td>10m28134f801</td>
<td>10m2814f4801</td>
<td>Raster</td>
<td>20120124</td>
<td>bbw_00_0</td>
</tr>
<tr>
<td>1</td>
<td>10m28134f801</td>
<td>10m2814f4801</td>
<td>Raster</td>
<td>20120124</td>
<td>bbw_00_0</td>
</tr>
<tr>
<td>108</td>
<td>10m28134f801</td>
<td>10m2814f4801</td>
<td>Raster</td>
<td>20120124</td>
<td>bbw_00_0</td>
</tr>
<tr>
<td>16</td>
<td>10m28134f801</td>
<td>10m2814f4801</td>
<td>Raster</td>
<td>20120124</td>
<td>bbw_00_0</td>
</tr>
<tr>
<td>12</td>
<td>10m28134f801</td>
<td>10m2814f4801</td>
<td>Raster</td>
<td>20120124</td>
<td>bbw_00_0</td>
</tr>
<tr>
<td>1</td>
<td>10m28134f801</td>
<td>10m2814f4801</td>
<td>Raster</td>
<td>20120124</td>
<td>bbw_00_0</td>
</tr>
<tr>
<td>1</td>
<td>10m28134f801</td>
<td>10m2814f4801</td>
<td>Raster</td>
<td>20120124</td>
<td>bbw_00_0</td>
</tr>
<tr>
<td>1</td>
<td>10m28134f801</td>
<td>10m2814f4801</td>
<td>Raster</td>
<td>20120124</td>
<td>bbw_00_0</td>
</tr>
<tr>
<td>1</td>
<td>10m28134f801</td>
<td>10m2814f4801</td>
<td>Raster</td>
<td>20120124</td>
<td>bbw_00_0</td>
</tr>
<tr>
<td>1</td>
<td>10m28134f801</td>
<td>10m2814f4801</td>
<td>Raster</td>
<td>20120124</td>
<td>bbw_00_0</td>
</tr>
<tr>
<td>1</td>
<td>10m28134f801</td>
<td>10m2814f4801</td>
<td>Raster</td>
<td>20120124</td>
<td>bbw_00_0</td>
</tr>
<tr>
<td>4</td>
<td>10m28134f801</td>
<td>10m2814f4801</td>
<td>Raster</td>
<td>20120124</td>
<td>bbw_00_0</td>
</tr>
<tr>
<td>4</td>
<td>10m28134f801</td>
<td>10m2814f4801</td>
<td>Raster</td>
<td>20120124</td>
<td>bbw_00_0</td>
</tr>
<tr>
<td>3</td>
<td>10m28134f801</td>
<td>10m2814f4801</td>
<td>Raster</td>
<td>20120124</td>
<td>bbw_00_0</td>
</tr>
</tbody>
</table>

References


© 2017 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).