Title of Paper: The Use of Mobile Positioning Data to Obtain Accommodation Statistics: Case Study of Indonesia

Abstract

The government of Indonesia has been set boosting the tourism sector as one of the targets to increase national income that is similar with the SDG’s goal: promote sustained, inclusive and sustainable economic growth, full and productive employment and decent work for all. A huge number of domestic tourists will accelerate local business in the destination area of the tourism especially for the accommodation sector. Statistics of accommodation that is used to support composing National Accounts is still obtained from a conventional survey. The issues of calculating the statistics from the survey include relatively low response rates, two months of time lag and a reluctance of respondents to answer due to repeated samples.

Currently, BPS as the National Statistics Office in Indonesia, tries to cover low response rates and uncompleted data using another similar conventional survey. Still, the result is not promising yet. The lack of integration of the conventional surveys generate inefficient repeated samples. BPS has therefore implemented the use of big data from mobile phone data to obtain complementary data about the tourism sector. Mobile Positioning Data (MPD) can be used to monitor people's movement and stays in the observed area. Joining-up the coordinate of the accommodation facility and data of MPD can predict the length of stay of tourists quite accurately. However, there are limitations using MPD, for instance expenditure and the number of hotel rooms booked cannot be covered. This paper aims to explain how the big data can promote a better solution for accommodation statistics.
I. Contents

I. Contents 2
II. Introduction 3
III. Overview of Accommodation Statistics 5
   A. Coverage 5
   B. Current Practice 6
IV. Exploring Big Data Source 8
   A. Fetching Location 8
   B. Mobile Positioning Data 10
      B.1. Accessing Data 11
      B.2. Data Structure 11
      B.3. Algorithm 13
      B.4. Technology 14
V. Issues and Challenges 16
VI. Conclusion 17
VII. References 18
II. Introduction

In 2015, world leaders agreed upon 17 Sustainable Development Goals (SDGs) to guide world development until 2030. Tourism is committed to do its part in this common endeavor. It is included as a target in three out of the 17 SDGs. They are “decent work and economic growth” (goal 8.9), “responsible consumption and production” (goal 12.b), and “life below water” (goal 14.7). Based on [1], tourism sector accounts for 10% of world GDP, 7% of global trade and one in ten jobs, where these achievements can contribute to all 17 goals. In Indonesia, tourism development is directed to become a reliable sector that capable in driving other economic sectors that are closely related. Ministry of National Development Planning (Bappenas) has set a tourism target for 2015-2019, where it stated that the target of tourism contribution to National GDP is 8%. To achieve this, of course, an increase in quality of tourism industry is needed.

Tourism industries that play a major role in tourism development includes accommodation services. The accommodation/hospitality industry is increasingly important, not only to increase tourists’ comfort, but also for economic impacts that will arise such as increasing income, expanding employment opportunities around them, and business opportunities. Based on these, it is very important to produce good accommodation statistics, which we currently achieved by conducting a survey of accommodation establishments.

For many countries, surveys of accommodation establishments are the most important short-term information source as they are in general, rather quickly available [2]. In more detailed regional details, accommodation statistics are most often the only source of information about tourism flows [2] [3]. To illustrate tourism flow, the number of arrivals and nights spent is the most widely used indicator. Of these two, nights spent are more appropriate to reflect the performance of the accommodation industry and the impact of tourist visits to places visited, as this indicator takes into account the full effect of the duration of the stay. Dividing the number of nights spent by the number of arrivals gives the average length of stay,
which can be used as an analytical indicator to offer additional information about the type of tourism in a country or region.

Currently to produce the average length of stay, Statistics of Indonesia (BPS) conducts enumerations/surveys every year. This enumeration is carried out in all regions of Indonesia which includes all commercial accommodation businesses, namely all-star hotels (1 to 5-star hotel), non-star/budget hotels, home stays, youth hostels, villas and other accommodation services such as bungalows and cottages [4]. However, there are several issues, one of them is non-response. Some of those accommodation establishments are not willing to provide their data. Therefore, other data sources can be an alternative to produce these accommodation statistics, for example the using of Big Data sources.

Big data is considered as part of data revolution, which can contribute to improve some aspects of statistical quality, such as timeliness and completeness, without compromising the relevance, impartiality, and methodological soundness of statistics [5]. Several National Statistics Office (NSO) have studied and utilized big data to produce official statistics such as online price utilization for the calculation of Consumer Price Index (CPI) by Singapore Department of Statistics [6], Statistics Canada did a research on feasibility of replacing and/or completing residential electricity consumption survey with smart meter data [7], daytime population estimations by Statistics Netherlands using mobile positioning data (MPD) [8], and many other utilizations.

Mobile positioning data is one of the big data sources that can be used in various applications [9]. Nowadays, almost everyone has a mobile phone to communicate, access information, mobile banking transactions, and many other activities. Every activity that uses mobile phone is actually recorded by mobile network operator (MNO) and eventually generate large amounts of data. This paper will examine the application of big data, especially mobile positioning data to produce accommodation statistics, namely to calculate the average length of stay. We also use Google Map to complete the calculation variable.
The paper is organized as follows. Section III is divided into two parts, they are; (1) Coverage, which gives an explanation regarding accommodation statistics, as well as the variables included in calculating the average length of stay; (2) Current practice, provides an overview of how BPS conducts surveys related to accommodation statistics and how information on average length of stay can be obtained. Section IV, explains the exploration of big data sources that will be used, as well as giving a map of big data sources with conventional survey variables. The last section provides conclusions of this study.

III. Overview of Accommodation Statistics

A. Coverage

International Recommendation of Tourism Statistics (IRTS) by United Nations (UN) suggest some indicators that can be obtained for completing accommodation statistics that should be performed regularly by the National Statistics Office (NSO). The indicators are room occupancy rates (gross or net), bed-place occupancy rates (gross or net), average number of persons per room, average room rate, average length of stay, average revenue per room night, average revenue per guest night, (average) revenue per available room (REVPAR), employees per room, average wage per employee, and revenue per employee [2]. IRTS also provides a mechanism to obtain data of market or commercial accommodation and non-market or non-commercial accommodation to enrich the statistics.

BPS adopted the recommendations from IRTS and other guidelines related to accommodation statistics that is officially issued by UNWTO and internationally comparable with other countries. Along with the purpose of constructing accommodation statistics, BPS conducts Survey of (Commercial) Accommodation monthly for directory sample (VHTS) and yearly for complete directory (VHTL). The monthly survey produces data as a fuel for constructing accommodation indicators
such as occupancy rate and average length of stay publication every month. The yearly survey is aimed to update commercial accommodation directory to be used in the next year monthly survey. Although not all indicators are produced, at least those two important indicators can be obtained to construct Tourism Satellite Account (TSA).

The occupancy rate is measured from the result of a monthly survey of commercial accommodation. Following IRTS definition, the indicator shows how many rooms have been sold during the month expressed as a percentage of the number of rooms available (or of the total number of existing rooms) during that same month. If all of an establishment’s rooms have been sold for every night of the month, the room occupancy rate is 100 percent. If only half have been sold, the rate is 50 percent. The average length of stay is counted using variables of the number of guests yesterday, number of today’s check-in and today’s check out from the monthly survey. The indicator is expressed in number of days.

The survey is held in all provinces covering all short-term commercial accommodation such as star hotels, non-star hotels, homestay, youth hostel, villa, bungalow, and cottage. Since BPS has representative offices in the regency level, the enumeration can be performed by BPS employees with sample at regency level.

**B. Current Practice**

Data compounding the average length of stay should be captured by exploiting potential data source that hypothetically more accurate and precise. The current process still relies on manual process of enumeration. Using Paper Assisted Personal Interviewing (PAPI) method, the enumerator asks number of available rooms (r), beds (b), last day used room (lr), today check-in’s rooms (cir), today check out’s rooms (cor), last day guests (lg), today check in’s guests (cig), and today check out’s guests (cog)
for domestic (d) and foreign guest (f). Focusing on the average length of stay, the indicator is measured from those variables using this formula:

\[
\begin{align*}
  fn_g &= lg_f + cig_f - cog_f \\
  dn_g &= lg_d + cig_d - cog_d \\
  avg_{los\_foreign} &= fn_g / cig_f \\
  avg_{los\_domestic} &= dn_g / cig_d \\
  avg_{los\_total} &= avg_{los\_foreign} / avg_{los\_domestic} \\
  avg_{los\_total} &= (fn_g + dn_g) / (cig_f + cig_d)
\end{align*}
\] ...(1) ...(2) ...(3) ...(4) ...(5) ...(6)

Data from all provinces are inputted using program entry which is expected able to be processed into tables and calculated produce average of length of stay automatically. However, we face an unstable response rate and inaccurate collected data that sometimes caused by human error either from respondent or enumerator. However, the program still can extract the data into an excel table which then calculated manually. The issues of post-processing data can be solved if the pre-processing data has been completed and firmed.

Facing low response rate, unanswered some important questions, inconsistency in contents, and impropriety variable of used rooms and number of guests forced data analyst to implement adjustment to fix the problems. Inappropriate adjustment process definitely increases bias on the indicators. In the specific conditions by some reasons, adjustment cannot be applied. As the result, the indicators for problematic provinces remain unpublished in the monthly official publication release. However, to state aggregate value of national level, BPS uses estimation.
IV. Exploring Big Data Source

A. Fetching Location

Location of commercial accommodation is usually stated by address, post code, and telephone number of the accommodation building and collected by enumerator. In line with developing of data collecting and the future purpose of data analysis, the survey collects coordinate location of commercial accommodation by copying latitude and longitude information from several coordinate fetcher mobile applications. Technically when the enumerator visits commercial accommodation for data collecting surveys, the enumerator runs the mobile app to get coordinate location then copying the information into the questionnaire.

But the problem is the surveys do not cover all directory of commercial accommodation yet, as the result, there are many hotels and other commercial accommodation do not have coordinate location in the database. We tried to get the information by browsing Google Maps to search the accommodation then copying the coordinate location information. The process runs one by one for around 28 thousand records and inefficient.

In order to get an alternative solution for the inefficient process, we build a robot using Kofax Kapow software. The robot tasks are to catch information of the commercial accommodation field by field and record by record from the database then combine the information into a complete directory. Then the robot will browse Google Maps website, placing the complete information into Google Maps’ search box then execute. Google Maps will show the information of coordinate location in the address link. The robot will fetch the coordinate location information into variable of longitude and latitude then write it into the selected record in the database. The process continues for the next iteration until the last record.
Image 1. The Robot to fetch coordinate location information

<table>
<thead>
<tr>
<th>A</th>
<th>id_iator</th>
<th>nama_iator</th>
<th>jenj_iator</th>
<th>kota_iator</th>
<th>koord_iator</th>
<th>lintang_biru_besar</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>20180304100</td>
<td>MS-HOTEL</td>
<td>BANDUNG</td>
<td>JL CIANGKRAN RANCABLUJU KM 08 KP SWAPULU</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>20180304101</td>
<td>RULIA MARIA PANGKAN</td>
<td>BANDUNG</td>
<td>JL RIAWA CIANGKRAN - PATENGKAN KM 3.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>20180304103</td>
<td>ORIANT VILLA</td>
<td>BANDUNG</td>
<td>JL SIDANG TRIP</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>20180304105</td>
<td>PATENGKAN VILLA</td>
<td>BANDUNG</td>
<td>JALU PATENGKAN</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>20180304107</td>
<td>PATIKA RESIDEN</td>
<td>BANDUNG</td>
<td>JL RIAWA CIANGKRAN PATENGKAN KM 9 JALU ENOKH PANCA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>20180304108</td>
<td>PEYSA WAHAN VILLA (RAHAYU)</td>
<td>BANDUNG</td>
<td>KP BEDANGKUNING RT 03 RW 05, DES, CEDEK, REI, CANGKUL</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>20180304111</td>
<td>ACEVI MAMA</td>
<td>BANDUNG</td>
<td>JALU KERTAMANAH NI 300 BP. GURUAN RT GURUAN DS. MAKAP</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>20180304111</td>
<td>PESONA NIRWANA WATERFARK BICOUGH</td>
<td>BANDUNG</td>
<td>KP CIANGKRAN DS PRAMAYUK KEC SOERANG</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>20180304122</td>
<td>PONDOK BENDA</td>
<td>BANDUNG</td>
<td>KP BARUTUNGAU DS. ALAM ENOKH RT 06 RW 06</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>20180304113</td>
<td>PONDOK NAHDIYA</td>
<td>BANDUNG</td>
<td>KP BATUUNGAU RT 06/01</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>20180304113</td>
<td>PONDOK NAHDIYA</td>
<td>BANDUNG</td>
<td>JALU TAMANAN PANCA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>20180304115</td>
<td>PONDOK SELLY</td>
<td>BANDUNG</td>
<td>JL RIAWA CIANGKRAN KM, NO 420</td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>20180304118</td>
<td>PONDOK SELLY 2</td>
<td>BANDUNG</td>
<td>PATENGKAN</td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>20180304119</td>
<td>PONDOK SEWA ALIANA</td>
<td>BANDUNG</td>
<td>KP BARUTUNGAU RT 03/01</td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>20180304112</td>
<td>JOROR VILLA</td>
<td>BANDUNG</td>
<td>KP BATOBAURO</td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>20180304120</td>
<td>PSI ENI VILLA</td>
<td>BANDUNG</td>
<td>KP KUJEDE</td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>20180304122</td>
<td>PUNJANG DIXTA</td>
<td>BANDUNG</td>
<td>JL SUNGUNG PUNTONS KOM 52</td>
<td></td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>20180304125</td>
<td>PURI PANJANGAN hotel</td>
<td>BANDUNG</td>
<td>JL RIAWA PANJANGAN NO 341</td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>20180304128</td>
<td>ROSA VILLA</td>
<td>BANDUNG</td>
<td>JALU RANCABLUJU PATENGKAN</td>
<td></td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>20180304129</td>
<td>PUNJANG DIXTA</td>
<td>BANDUNG</td>
<td>JL RIAWA GULF NO 22 DUGO, RT 01 RW 06, DESA CIURUL</td>
<td></td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>20180304131</td>
<td>SAYINDA</td>
<td>BANDUNG</td>
<td>KEC CIANGKRAN</td>
<td></td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>20180304134</td>
<td>SAYING PRASANGI</td>
<td>BANDUNG</td>
<td>JL RIAWA PANJANGAN</td>
<td></td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>20180304153</td>
<td>SAUSA GUNCUT KAP-VILLA</td>
<td>BANDUNG</td>
<td>JALU CIANGKRAN WIATAU AMPATU, PATENGKAN</td>
<td></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>20180304154</td>
<td>SAUSA GUNCUT KAP-VILLA</td>
<td>BANDUNG</td>
<td>JALU KERTAMANAH WIATAU AMPATU, PATENGKAN</td>
<td></td>
<td></td>
</tr>
<tr>
<td>26</td>
<td>20180304138</td>
<td>SAUSA BULU (TED) HOTEL</td>
<td>BANDUNG</td>
<td>KP PANCA SAPRI</td>
<td></td>
<td></td>
</tr>
<tr>
<td>27</td>
<td>20180304161</td>
<td>ASI ENE MONEY ORANI</td>
<td>BANDUNG</td>
<td>CIANGKRAN</td>
<td></td>
<td></td>
</tr>
<tr>
<td>28</td>
<td>20180304140</td>
<td>SAUSA GUNP</td>
<td>BANDUNG</td>
<td>JL RIAWA CIANGKRAN - RANCABLUJU KM, 6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>29</td>
<td>20180304141</td>
<td>SAUSA KATUNANG KAWINAY</td>
<td>BANDUNG</td>
<td>CIANGKRAN</td>
<td></td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>20180304163</td>
<td>SAUSA GUNP</td>
<td>BANDUNG</td>
<td>JL SENG PT 01 / K 18 DI PANJANGAN</td>
<td></td>
<td></td>
</tr>
<tr>
<td>31</td>
<td>20180304164</td>
<td>SCARLET HOTEL INDONESIA</td>
<td>BANDUNG</td>
<td>JALU RIAWA PANJANGAN TULU 1 NO 15A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>20180304165</td>
<td>SEDAYUNA PENGINAPAN</td>
<td>BANDUNG</td>
<td>JALU BAKANGSAU NO 345 CIANGKRAN</td>
<td></td>
<td></td>
</tr>
<tr>
<td>33</td>
<td>20180304147</td>
<td>SENASA VILLA (MARIO)</td>
<td>BANDUNG</td>
<td>KP SUKASARI DESA NISALARI</td>
<td></td>
<td></td>
</tr>
<tr>
<td>34</td>
<td>20180304148</td>
<td>SHINTA CORNER RANCH AND RESORT VILLA</td>
<td>BANDUNG</td>
<td>JL RIAWA SURAHMEMIA RT 02 RW 13 NO 23 PANJANGAN</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Image 2. Initial Database of Commercial Accomodation
This final database will be used as a baseline for determining whether a tourist uses accommodation services or not. This identification process clearly crucial before moving forward into next step in predicting the number of guests in the area of accommodation.

### B. Mobile Positioning Data

Mobile positioning is the tracing of the location coordinates of mobile phones [10]. Mobile positioning can be divided into active and passive positioning [11]. Active mobile positioning is used for mobile tracking in which the location of the mobile phone is determined (asked) with a special query using a radio wave. In order to track certain phones for research projects, a special permit from the phone holder is required.

Passive mobile positioning data are location data stored automatically in the database of mobile operators when a person uses a mobile phone. This paper will focus on exploring passive mobile positioning.
B.1. Accessing Data

In this era, mobile phone usage has become a lifestyle not only for corporate use but also personal use that cause increasing number of mobile phone user. Any activities of the mobile phone are recorded in Mobile Network Operator’s (MNO) database that provide accessibility of data. This rise the opportunity for researcher to get more accurate and real-time data by extracting from the mobile phone use data from MNO [13][14].

In 2018, BPS collaborated with Telkomsel, one of the largest Mobile Network Operators (MNO) in Indonesia for exploring MPD. Telkomsel provides data access to BPS through sandboxing in Telkomsel office. The Sandbox contains data samples of 15 thousand Telkomsel subscribers for exploration and 65 subscribers who are BPS volunteers to for data validation. The currently available results are domestic tourism statistics and mobility statistics, namely commuter and circular statistics. Based on the results of the exploration carried out last year, this paper will review and prepare for the next development of MPD for accommodation statistics.

B.2. Data Structure

To properly use the raw data of mobile phone usage extracted from MNO’s database, we need effort in translating our analysis purposes into suitable methods and algorithms. In line with the requirement, passive mobile positioning method is suited to the purpose. The method may come from customer transaction data or Call Data Record (CDR), and customer presence location or Location Based Service (LBS). Every paid transaction made by cellular customers will form CDR data, while LBS is formed periodically due to the presence of cellular customers at a certain location.

There are two main tables in the passive mobile positioning database that can be utilized to generate accommodation statistics, especially the calculation of average...
length of stay. The first table contains activity data of cellular customer data that normally collected to the precision of one network cell. The second table is an area reference for each transaction recorded in the first table. In this paper the first table is then referred as the raw_data table, while the second table is region_reference.

Following are the list of fields/variables contained in each table and their description:

1. raw_data, consists of:
   - hashed_msisdn: hashed cellular customer number
   - trx_date: transaction date
   - datetime: transaction time (date, hour)
   - lac: Base Transceiver Station (BTS) area code
   - cell_id: BTS cell identity
   - calltype: transaction type, consists of CHG_POST (recorded if there is a call/Short Message Service (SMS) out), CHG_PRE (recorded if there is a call/ SMS in), LBA_ALL (recorded every there is a location update from where the cellular customers located).
   - timezone: time zone, consists of WIB, WITA, and WIT.

2. region_reference, consists of:
   - lac: Base Transceiver Station (BTS) area code
   - cell_id: BTS cell identity
   - province: geographical area in province level
   - districts: geographical area in districts level
   - sub-district: geographical area in sub-district level
   - kelurahan: geographical area in kelurahan level
   - longitude: longitude of BTS
   - latitude: latitude of BTS
The two tables will be joined based on lac and cell_id to get spatial information. Each record that recorded in raw_data table will be complemented by geographical description based on region_reference table.

B.3. Algorithm

The algorithm of average length of stay calculation using mobile positioning data is as follows:

1. Identify each trip on each cellular customer that is classified as a tourism trip.

To identify tourism trips, the steps are:

   i. Home, office, and usual environment identification of each cellular customer. Several algorithms for determining the usual environment have been developed [10] [12] [13].

   ii. Discarding the trip detected as the usual environment.

   iii. The remaining trip is a trip which is likely a tourism trip.

   iv. Applying filtering to the remaining trips to be considered as a tourism trip. For example, a trip that is considered as tourism trip is if the cellular customer is staying in a place for a minimum of 6 hours. This filtering can differ between countries that will implement the algorithm.

2. Identify tourism trips that at night (eg. 20.00 - 06.00) are staying around the coordinates of the hotel that have been previously obtained. For example, \( r \) is a radius that is set to widen the precision of determining someone to be in a hotel, then a tourism trip that is considered to use accommodation services is a trip that has coordinates in the range of coordinates of the hotel + \( r \).

3. For each date:
i. Calculates last day guests (lg), which is from each trip found in stage 2 and then identifies a trip detected in coordinates around the hotel above 2:00 p.m. on the day before the corresponding date, and is not detected in those coordinates on the previous two days. The specified hour limit can differ from one country to another, depending on the characteristics constraint in that country.

ii. Calculate today check in’s guests (cig), which is from each trip found in stage 2 and then identify the trip detected in coordinates around the hotel above 02.00 p.m. on the day of the corresponding date and not detected at that coordinate in previous day.

iii. Calculate today check out’s guests (cog), which is from each trip found in stage 2 and then identify the last detected trip in the coordinates around the hotel under 12.00 a.m on the day of the corresponding date and undetected around the hotel coordinates the next day.

iv. Calculate the number of nights that are used based on formulas (1) and (2).

v. Calculate the average length of stay based on formula (3). The number of guests is obtained from unique hashed_msisdn which is detected on stage (2).

B.4. Technology

Mobile positioning data with a very large size is stored in a repository that is based on Hadoop technology. To process the data, we must use programming language and scripting that supports it. Some of them are scripting using HiveQL, as well as the use of programming languages based on Map Reduce. The risks that must
be faced by NSO in using MPD are special skills that must be possessed by employees in big data processing as mentioned.

C. Mapping

The following is a mapping between variables obtained in conventional survey compared to the use of big data:

<table>
<thead>
<tr>
<th>No.</th>
<th>Conventional Survey</th>
<th>Big Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Hotel directory</td>
<td>Google Map</td>
</tr>
<tr>
<td>2.</td>
<td>Last day guests (lg)</td>
<td>Algorithm step 3.i</td>
</tr>
<tr>
<td>3.</td>
<td>Check in’s guests (cig)</td>
<td>Algorithm step 3.ii</td>
</tr>
<tr>
<td>4.</td>
<td>Today check out’s guests (cog)</td>
<td>Algorithm step 3.iii</td>
</tr>
<tr>
<td>5.</td>
<td>Number of guests</td>
<td>Unique hashed_msisdn</td>
</tr>
</tbody>
</table>

D. Data Quality and Future Plan

Data quality with radius has been proven with Asian Games (AG) project that we did last year with Bappenas, Telkomsel, and Positium. Similar to our approach with accommodation statistics calculation above, we have used subscribers stays in a certain location radius with mobile positioning data to predict the number of AG visitors in the venues which was held in Jakarta and Palembang on August 2018. Our calculation results can be seen in [15], [16], and [17]. We compared our AG visitor calculation with MPD against other supporting data such as visitor/venue ticket sales and we are confidence with the result.

To use this method to calculate accommodation statistics using MPD with the MNO, we need to establish new agreement with the MNO to include accommodation statistics calculation as one of the MPD results in this year’s procurement. We plan to work on this with the MNO this year. As we did last year with the domestic tourism calculation with MPD, we will iterate and validate the algorithm until we are confidence with the result.
V. Issues and Challenges

The new approach of data collecting for official statistics still continue to be developed. The response rate issue is one of the biggest problems that needs to be solved. The possible solution to increase the response rate is using a technological approach into one or more procedures of survey. Even though mobile positioning data as one of technological approaches provides promising solutions to be applied, the difficult situations cannot be avoided.

Traditional and old-fashioned mindset that views mobile positioning data is not suitable for changing traditional usual process are the main problem of MPD development. Their arguments are the questionable validity of the methodology and the nil interaction with the respondents will reduce the potential data verification process. However, MPD has been proven to catch real-time movement data of people accurately that can be used to complement traditional survey’s data because not all people can recall their long-term memory to answer every question of the traditional survey.

There are data of some variables that can be caught accurately using MPD e.g. mobile user location and movement in every second precisely. It is more appropriate to be used rather than forcing respondents to remember their activities that has been done in the past six months or one year ago. However, there are several indicators that cannot be generated from MPD, such as the number of rooms or the number of beds. Therefore, mixed-mode/multimode data collection is the best solution, for example, by combining conventional surveys with big data sources such as mobile phone data.

As one of important players in the MPD development, MNO is critical to be a participant of joint partnership with NSO. The regulation of the procurement process tightly binds both NSO and MNO where MNO as the data provider and NSO as business partner. The process of making the agreement has to deal with privacy issues of mobile data usage to protect both sides from a lawsuit in the future.
On the other hand, MPD as a new thing in the current decade still need to be developed and firmed. The algorithm is not mature yet because each location or country has their own problematic conditions such as geographical issues, unique mobile phone user behaviors, privacy issues, etc. Those problems challenge researchers to update and generate their approaches into a suitable algorithm that can be applied in many regions.

Currently, we are only confidence with MPD results at regency level. In order to increase the research confidence and credibility of the methodology, MPD method should be taken to the next level estimation that can predict movements in smaller area i.e. kecamatan (sub-district) with satisfying degree of confidence. Increasing the level of estimation with the assumption all needed factors are held and firmed will raise acceptance possibility from decision makers. One of the issues that need to be tackled to achieve that is how to correctly predict subscriber’s location in the sub-district border areas.

VI. Conclusion

The current conventional survey of commercial accommodation is performed monthly by direct interviewing the representative of commercial accommodation. Even though the value can be adjusted using the phenomenon data and other supporting data, the problem of low response rate, unanswered some important questions, inconsistency in contents, and impropriety variable of used rooms and number of guests have to be reduced. Big data has potential advantages as an alternative data source to complete the blanks.

Mobile positioning data has been proven to track user movement in the such area during a certain time of analyzing. By matching coordinate location of commercial accommodation and coordinate location of mobile phone user, we can extract the number of mobile phone user as a guest in certain night in the commercial accommodation. Knowing the promising methodology to produce alternative data to construct the indicators, we suggest to use the methodology of mobile positioning data.
VII. References


