The surge in the usage of mobile devices and Internet services is generating an enormous amount of data, which has a high and largely untapped potential to support the discovery of new knowledge about human behaviors. The data that can be collected in production mobile networks is already proving an invaluable proxy to analyze the habits of large populations at large scales of cities or countries, complementing and in some cases replacing traditional sources such as surveys or censuses that are expensive and time-consuming to run. Examples of the substantial utility of mobile network data for research abound, and span a plethora of domains: the data can unlock analyses of mobility patterns [1–5] and social interactions [6], explorations of transportation systems [7] estimates of static and dynamic population density [8–11], predictions of poverty [12, 13], socioeconomic inequality [14, 15] or digital divides [16], and mappings of land usage [17–20] or urban transformation [21] or pollution [22]; in addition, mobile network data have proven instrumental to assess the impacts of natural disasters [23] or infectious disease transmission [24–26], and the effectiveness of the associated containment policies [27, 28].

These data can also enable studies aimed at understanding how the mobile network infrastructure is used, improving its management and extend its functionalities: for instance, they can be leveraged for the localization and tracking of devices [29, 30], the characterization of network loads [31, 32] and application usages [33], the prediction of traffic fluctuations [34], the planning [35] and dynamic reconfiguration [36] of Radio Access Networks (RAN) infrastructures, or the data-driven management of network resources in sliced environments [37]. The list above only covers a small sample of the literature and does not pretend to be comprehensive by any means; surveys are available that provide a more thorough review of the relevant body of works [38, 39].

Despite their prospective impact on a wide range of domains, mobile network data are hard to come by. The sensitivity of the information they provide, the concerns for the privacy of the data subjects, and the questions on the advantage that they could provide to market competitors are all reasons why mobile network operators and service providers typically regard the data as confidential, and are not prone to share them with the research community. It limits access to mobile network data, curbing innovation as well as preventing verifiability and reproducibility of the research results whenever permission to use some data is granted under restrictive Non-Disclosure Agreements (NDAs).

In this paper, we present a novel dataset of mobile data traffic that is open to the research community within the context of a challenge organized jointly with the NetMob 2023 conference¹. By doing so, we revive a recent tradition of challenges based on mobile phone data that have sparked substantial interest and a flurry of original research outcomes a few years ago. The Data for Development (D4D) challenges organized by Orange in collaboration with NetMob in

¹NetMob is the primary conference on the analysis of mobile network data to investigate social, urban, societal and industrial problems. Details can be found at https://netmob.org/.

ABSTRACT
Digital sources have been enabling unprecedented data-driven and large-scale investigations across a wide range of domains, including demography, sociology, geography, urbanism, criminology, and engineering. A major barrier to innovation is represented by the limited availability of dependable digital datasets, especially in the context of data gathered by mobile network operators or service providers, due to concerns about user privacy and industrial competition. The resulting lack of reference datasets curbs the production of new research methods and results, and prevents verifiability and reproducibility of research outcomes. The NetMob23 dataset offers a rare opportunity to the multidisciplinary research community to access rich data about the spatio-temporal consumption of mobile applications in a developed country. The generation process of the dataset sets a new quality standard, leading to information about the demands generated by 68 popular mobile services, geo-referenced at a high resolution of 100 × 100 m² over 20 metropolitan areas in France, and monitored during 77 consecutive days in 2019.

1 INTRODUCTION
The surge in the usage of mobile devices and Internet services is generating an enormous amount of data, which has a high and largely untapped potential to support the discovery of new knowledge about human behaviors. The data that can be collected in production mobile networks is already proving an invaluable proxy to analyze the habits of large populations at large scales of cities or countries, complementing and in some cases replacing traditional sources such as surveys or censuses that are expensive and time-consuming to run. Examples of the substantial utility of mobile network data for research abound, and span a plethora of domains: the data can unlock analyses of mobility patterns [1–5] and social interactions [6], explorations of transportation systems [7] estimates of static and dynamic population density [8–11], predictions of poverty [12, 13], socioeconomic inequality [14, 15] or digital divides [16], and mappings of
2013 and 2014 [40, 41], the International Telecommunication Union (ITU) challenge to investigate the Ebola epidemic in West African countries in 2015 [42], the Telecom Italia Big Data Challenge launched in 2014 and 2015 [43], the Data for Refugees (D4R) challenge conducted by Turk Telekom in 2018 [44] or the Future Cities Challenge supported by Foursquare during the NetMob 2019 conference [45] are prominent examples of such past initiatives.

Our dataset, referred to as NetMob23 hereinafter, sets a new standard for mobile network data made available to the research community, from multiple perspectives as follows.

- While previous datasets have largely focused on Call Detail Records (CDRs) that only capture network events associated with voice calls and text messages that are sparse and irregular over time, NetMob23 contains information about the data traffic generated by the mobile devices attached to a modern 4G cellular network, which has been for the past ten years the vastly predominant way of accessing wireless network services.
- The NetMob23 dataset captures traffic in a developed country like France, which offers a different perspective than earlier datasets focusing on developing countries; also, the data spans 20 metropolitan areas in France, offering the possibility of generalizing analyses and juxtaposing results across heterogeneous urbanization levels and population densities.
- Unlike any dataset previously available to the research community within the framework of open challenges, NetMob23 offers rich information about the usage of 68 popular mobile services, which opens significant opportunities to understand the consumption of applications and its implications across research domains.
- The original generation process behind the NetMob23 dataset makes a major step beyond the legacy approach of using Voronoi tessellations as a proxy for antenna coverage, and results in a dataset of unprecedented spatial accuracy where the mobile data traffic information is mapped to more than 870,000 high-resolution regular grids whose individual elements span $100 \times 100$ m$^2$ each, for a total of over 440 billion data points.

Overall, the NetMob23 dataset has the potential to support many novel explorations of mobile network traffic and enable the development of new applications on top of those findings. We look forward to seeing creative and constructive uses of the data by the research community worldwide.

## 2 DATA SOURCES

The generation process underpinning NetMob23 hinges upon open-source geospatial data and extensive measurements from Orange, a major mobile network operator in France. We note that Orange roughly has a 30% market penetration in the country that is relatively uniform across the French territory. This provides a solid statistical basis for downstream analyses that generalize to the entirety of the local population. We next present the different data sources, and discuss the ethical standards of the data collection and processing.

### 2.1 Metropolitan areas

We use geographical shapefiles outlining the 20 major metropolitan regions in France, a zoning employed by national and local administrations to carry out joint planning of the educational, cultural, economic, and social initiatives over the country territory. Each metropolitan region encompasses a...
set of neighboring *communes*, which are French local administrative zones analogous to a civil township in the United States, and thus covers both dense urban, suburban, and more rural areas that constitute the conurbation of a specific city. Figure 1 shows the location and contour of all metropolitan areas covered by the NetMob23 dataset, overlapped to a country-wide map of population density. The areas of a few sample cities are magnified to appreciate the diversity of population density captured by each of them.

Overall, our dataset includes the foremost industrial, commercial, and financial centers of France, which are home to more than one third of the total population in the country.

### 2.2 Mobile network traffic

We employ mobile data traffic collected over the 20 target metropolitan areas for 77 consecutive days, *i.e.*, roughly two and a half months, from March 16, 2019, to May 31, 2019.

#### 2.2.1 Service-level traffic volumes

The traffic measurements were performed by Orange using passive measurement probes tapping at the Gi, SGi and Gn interfaces connecting the Gateway GPRS Support Nodes (GGSNs) and the Packet Data Network Gateways (PGWs) of the of Long Term Evolution (LTE) Evolved Packet Core (EPC) network to external public data networks (PDNs). This monitoring strategy allows capturing all 4G traffic traversing the mobile network serving the whole country. The probes run dedicated proprietary classifiers that allow associating individual traffic flows to their corresponding mobile applications to enable network monitoring, traffic engineering and research activities.

Ultimately, the EPC probes gather information about the uplink (UL) and downlink (DL) volume of the demand generated by each of 68 mobile services, which are together responsible for about 70% of the total mobile data traffic observed by Orange in France. Figure 2 offers a look into the applications included in the dataset, and a ranking of the same based on the fraction of total traffic they generate. We observe that the services responsible for the largest demands are responsible for around 10% of the overall mobile network usage. The logarithmic scale of the ordinate highlights the power law that characterizes the relative consumption across applications, and the consequent diversity of per-service traffic, which spans three orders of magnitude when juxtaposing the top and bottom applications in the dataset.

#### 2.2.2 Traffic flow to eNodeB association

We associate traffic volumes to specific base stations using network signalling data (NSD) captured by probes monitoring the LTE S1 interface connecting eNodeBs, *i.e.*, 4G base stations, to the Mobility Management Entity (MME). The NSD allow tracking the eNodeB serving a mobile device across all control-plane events, which include (i) voice and texting communications such as call establishments and SMS transmissions, (ii) handovers, *i.e.*, device cell changes during communication, (iii) Tracking Area (TA) updates, *i.e.*, cell changes that cross boundaries among larger regions that trigger control messages also from idle devices, (iv) active paging, *i.e.*, periodic requests to update the location of the device started from the network side, (v) network attaches and detaches generated by devices joining or leaving the network as they are turned on/off, or (vi) data connections, *i.e.*, requests to assign resources for traffic generated by mobile applications running on the device. Such high-frequency NSD events allow associating each traffic flow to the exact sequence of eNodeBs that serviced it, and thus to accurately link (portions of) the traffic volume of the flow to each base station.

#### 2.2.3 Service-level traffic time series at eNodeBs

We aggregate the traffic volume generated by all uplink or downlink flows pertaining to a given mobile service and served by a same eNodeB. We also aggregate the resulting traffic volumes over 15-minute time intervals, *i.e.*, a temporal granularity that allows observing a wide range of time-dependent phenomena, while keeping the dataset at a reasonable size.
Overall, the mobile network traffic data we employ is in the form of time series of the uplink and downlink load generated by 68 mobile applications at every eNodeB, with 15-minute time steps. In order not to disclose the sensitive information of the actual volume of traffic served by the mobile network operator, we normalize all traffic by a same random value. Therefore, traffic is not provided with a specific unit (e.g., bytes) but it is still fully comparable across space and time.

2.3 Coverage dataset

Having associated the correct service-level traffic to each eNodeB over time, we employ coverage information for each eNodeB to perform a spatial mapping of the traffic information to the geographical space. This allows producing data with a much higher spatial resolution than using legacy approaches such as Voronoi tessellations.

Specifically, we start from coverage data computed using a commercial radio-frequency signal propagation tool. For every eNodeB, coverage is encoded as probabilities of association over a regular grid tessellation with tiles of $100 \times 100$ m$^2$ each. A bi-dimensional matrix of $600 \times 600$ tiles is produced for a single eNodeB, hence providing complete coverage information over an area of $60$ km$^2$ surrounding the base station. The matrix tiles contain probabilities $p(i \mid ℓ)$ that explain the likelihood of a User Equipment (UE) to connect to eNodeB $i$ while at tile $ℓ$. By inverting this probability information via Bayes’ theorem, under the assumption of a uniform density of population within the area covered by the eNodeB, we derive $p(ℓ \mid i)$, i.e., the probability that a device associated with a base station $i$ is located at tile $ℓ$. We use $p(ℓ \mid i)$ to distribute over the high-resolution grid of $100 \times 100$-m$^2$ tiles the service-level traffic observed by each eNodeB in a probabilistic fashion, as detailed in Section 3.

3 GENERATION METHODOLOGY

The different data presented above are combined to generate the final NetMob23 dataset, following a process that is visually summarized in Figure 3. Formally, let us denote by $T_{a}^i(t)$ the mobile traffic generated by application $a$ at eNodeB $i \in I$ during time slot $t \in T$, where $I$ is the set of all base stations and $T$ denotes the whole system observation period. Also, recall that $P(ℓ \mid i)$ is the probability of a user to be at a location $ℓ$ while being connected to eNodeB $i$.

As portrayed in Figure 3, we first multiply traffic observed at each eNodeB (denoted by B in the figure) by the UE positioning probability of the same eNodeB (A in the figure), so as to distribute over space the traffic volume observed at the base station. This operation is repeated for each time slot $t$ of the time series of every mobile service (e.g., Spotify in the figure). The result is a traffic map $M_{a}^i(t)$, which represents how the traffic generated by a given service $a$ (again, Spotify in the figure example) at the target eNodeB $i$ is probabilistically distributed over the geographical space at time $t$.

Since the same regular grid tessellation is used for probabilities $P(ℓ \mid i)$ across all base stations $i \in I$, it is possible to repeat the process above for all eNodeBs, and generate consistent maps of the spatial traffic of a same application $a$ over the whole mobile network. This is attained by simply adding...
Table 1: Dimensions of the matrices of regular grid tiles in each of the 20 target urban regions.

<table>
<thead>
<tr>
<th>Region</th>
<th>Rows</th>
<th>Columns</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bordeaux</td>
<td>334</td>
<td>342</td>
</tr>
<tr>
<td>Clermont-Ferrand</td>
<td>208</td>
<td>268</td>
</tr>
<tr>
<td>Dijon</td>
<td>499</td>
<td>234</td>
</tr>
<tr>
<td>Grenoble</td>
<td>330</td>
<td>231</td>
</tr>
<tr>
<td>Lille</td>
<td>426</td>
<td>342</td>
</tr>
<tr>
<td>Lyon</td>
<td>228</td>
<td>287</td>
</tr>
<tr>
<td>Marseilles</td>
<td>211</td>
<td>246</td>
</tr>
<tr>
<td>Metz</td>
<td>226</td>
<td>210</td>
</tr>
<tr>
<td>Montpellier</td>
<td>334</td>
<td>327</td>
</tr>
<tr>
<td>Nancy</td>
<td>151</td>
<td>165</td>
</tr>
<tr>
<td>Nantes</td>
<td>277</td>
<td>425</td>
</tr>
<tr>
<td>Nice</td>
<td>150</td>
<td>214</td>
</tr>
<tr>
<td>Orleans</td>
<td>252</td>
<td>253</td>
</tr>
<tr>
<td>Paris</td>
<td>282</td>
<td>346</td>
</tr>
<tr>
<td>Rennes</td>
<td>409</td>
<td>370</td>
</tr>
<tr>
<td>Saint-Etienne</td>
<td>423</td>
<td>501</td>
</tr>
<tr>
<td>Strasbourg</td>
<td>305</td>
<td>251</td>
</tr>
<tr>
<td>Toulouse</td>
<td>296</td>
<td>238</td>
</tr>
<tr>
<td>Tours</td>
<td>280</td>
<td>347</td>
</tr>
</tbody>
</table>

the different traffic maps as $M_a(t) = \sum_{i=1}^{I} M_i^a(t)$ (shown in Figure 3 as C). The final $M_a(t)$ is the overall traffic map of service $a$ at time $t$, defined over locations $\ell$ (D in the figure).

We repeat the above steps for all $t \in T$ and for all applications to ultimately obtain a complete spatiotemporal representation of the service-level traffic in the 20 target metropolitan areas introduced in Section 2.1 over a high-resolution grid of locations $\ell$.

### 3.1 Ethics considerations

The mobile network traffic dataset we use to generate the dataset was collected, processed and aggregated as described in Section 2.2 in full compliance with Article 89 of the General Data Protection Regulation (GDPR) [46], under the supervision of the Data Protection Officer (DPO) at Orange. In particular, all data management was performed on a secure platform at the operator’s premises and the raw data was deleted immediately afterward.

The resulting service-level time series represent traffic aggregated over all UEs both in space, at eNodeB level, and time, over 15-minute intervals. Moreover, the traffic associated to different base stations is further aggregated via the spatial mapping described earlier. The final representation does not allow re-identifying or tracking individual users.

### 3.2 Final dataset format

To facilitate access to the data, the NetMob23 dataset is divided into spatial representations and traffic information, which are detailed next. As part of the material provided with the dataset, we also make available Jupyter notebooks in Python with examples of manipulations of the dataset and calculations of high-level statistical indicators and plots.

#### 3.2.1 Spatial representation

We publish a GeoJson file for each metropolitan region with the WGS84 coordinate system. Each of these files contains the tile identifier of each regular grid tile in the target urban area, and the corresponding polygon that bounds the $100 \times 100$-m² geographical surface of the tile. We also provide an alternative matrix representation of the space, in the form of a text file containing the number of rows and columns in the matrix, whose values are also shown in Table 1.

#### 3.2.2 Traffic information

The traffic dataset folder is organized according to a tree structure, as shown in Fig 4. The root is the target urban region, followed by the application name and then the day, with two leaves, i.e., text files, for the UL and DL directions, respectively. In each text file, one line represents the (normalized) traffic in a given tile of the city, for the corresponding application and date, from midnight to 23:45 at 15-minute time steps, as shown in Table 2.

More precisely, each line contains the following fields:

- **tile** is the tile identifier used in the GeoJson spatial representation file for the same urban area. In the case of the matrix representation of tiles, the position inside the matrix can be retrieved by an integer division and a modulo operation.
- **t hh:mm** is a vector of 96 values, corresponding to the traffic observed in UL or DL at location tile for the target city, service, and day, during all 15-minute time slots, starting from midnight to 23:45. Due to daylight

![Figure 4: Hierarchical organization of traffic files.](https://github.com/nds-group/netmob2023challenge)
<table>
<thead>
<tr>
<th>Period</th>
<th>Affected regions</th>
<th>Services</th>
</tr>
</thead>
<tbody>
<tr>
<td>April 9, 2019</td>
<td>Bordeaux</td>
<td>All</td>
</tr>
<tr>
<td>April 12, 2019</td>
<td>Bordeaux, Toulouse</td>
<td>All</td>
</tr>
<tr>
<td>April 14, 2019 [47]</td>
<td>Nationwide</td>
<td>Meta Inc.</td>
</tr>
<tr>
<td>May 12, 2019 [48]</td>
<td>Nationwide</td>
<td>All</td>
</tr>
<tr>
<td>May 22-25, 2019</td>
<td>Bordeaux, Toulouse</td>
<td>All</td>
</tr>
</tbody>
</table>

Table 2: Example of the format of a traffic file.

Table 3: Major anomalies detected in the dataset.

4 QUALITATIVE ANALYSIS

In this section, we provide a preliminary exploration of some characteristics of the NetMob23 dataset. The analysis aims at discussing basic properties of the spatial and temporal evolution of the service-level mobile traffic, and highlight anomalies inherent to the dataset that may bias downstream usages of the same.

4.1 Anomaly analysis

The NetMob23 dataset is based on traffic information measured in a production network. As such, it is characterized by anomalous events of technical nature, caused, e.g., by exceptional surges in demands, radio access network malfunctioning, network configuration errors, energy supply problems, or issues in the traffic monitoring system.

We report examples of anomalous events that can be observed in the NetMob23 dataset in Figure 5, for different combinations of cities and services. The plots show the overall traffic in UL and DL, with anomalies evidenced by light red areas. The three plots show for instance the effects of a nationwide network outage on May 12, 2019, which caused both UL and DL traffic to drop substantially in the whole Orange network. The top and bottom plots also highlight an event affecting only the South-East of France but for a longer period spanning from May 22 through May 25: the problem caused a reduction of traffic in Bordeaux and Toulouse, but not in Paris, due to the geographical locations of the cities. More rarely, outages concern specific applications, such as for the services of Meta on April 14, as seen, e.g., in the top plot for Facebook Live traffic. The full list of major anomalies we detected in the data is provided in Table 3.

4.2 Temporal Analysis

People tend to follow fairly regular patterns in their daily lives, and mobile network data have been repeatedly shown to be affected by such periodicity [1, 20, 31]. This is true at the level of the total mobile traffic, but even more so when considering individual applications, since users employ those at specific times of their weekly routine.

Figure 6 shows the time series of four major mobile applications with very diverse usage patterns in Paris. LinkedIn is a business and employment-oriented social media application that is used to connect with other professionals to share their resumes, work, or professional events; Netflix is an application that people use for entertainment; Apple iCloud is a platform that is used to store and sync data on various Apple devices for applications like Apple Mail, Apple Calendar, Apple Photos, Apple Notes, contacts, settings, and backup files; and, Uber is used for long and short rides. Each
Figure 6: Traffic time series of four different applications in Paris, as observed during a same week.

Figure 7: Traffic time series of the Netflix service in four different cities during a same week.
of these applications shows different patterns in their time series, i.e., LinkedIn shows a high traffic peak in the early morning hours until the afternoon and then the traffic values for both UL and UD start to decrease. Conversely, Netflix shows a high traffic peak in the late evening because people are usually free during these hours, while Apple iCloud does not follow either the day or night pattern. Finally, Uber generates a clear pattern where weekend traffic is almost the same during the day and night, implying that the service is used more on weekends when people travel and take rides to return late at night. As another notable difference across applications, the UL/DL ratio is very heterogeneous, which could be noted already in Figure 2.

These behaviors are not a specific feature of a large metropolis like Paris. A same application tends in fact to be mostly used in the same way in different geographic locations, as shown in Figure 7. There, we consider the example of Netflix, and display its usage in four cities, namely Bordeaux, Lyon, Toulouse, and Marseille. The patterns are very much comparable in all plots, although minor changes occur that would deserve a deeper investigation.

4.3 Spatial Analysis

The NetMob23 dataset allows investigating spatial properties of the mobile service usage as well. Figure 8 shows the average traffic maps on Mondays (top row) and Sundays (bottom row) for the same four applications considered before, i.e., Netflix, LinkedIn, Apple iCloud, and Uber. Each of the plots shows the whole Paris urban area and a zoom on the city center delimited by the local inner beltway.

The figure allows comparing the applications along two dimensions. On the one hand, looking at plots from left to right shows that the different applications have very diverse geographical distributions of their generated traffic. Netflix, for example, is pervasively distributed across the region, which is even more evident in the zoomed map, whereas LinkedIn and Apple iCloud have a high concentration in the part of the city where there are many large offices and workplaces. In the case of iCloud, notable traffic is also recorded at touristic spots, amusement parks, or entertainment facilities: we speculate that these may be locations where people usually upload photos or videos to their personal cloud. In the case of Uber, there is a clear distribution of traffic on city streets, highways, train stations, and airports.

On the other dimension, a comparison of the demands for a same application in working and weekend days, i.e., from top to bottom of Figure 8, also reveals interesting phenomena. For instance, Netflix is used in a fairly constant way during the week, whereas Uber is more actively used on weekends than on weekdays. On the contrary, LinkedIn and Apple iCloud largely drops in weekends, although Apple iCloud usage persists in the more touristic areas of the city.
5 ADDITIONAL RESOURCES

The NetMob23 dataset can be enriched via combination with other sources of information. We list a number of sources of sociodemographic and telecommunication indicators that cover the French territory at around the same time of the mobile traffic data collection.

Administrative boundaries. These resources are available on the Open platform for French public data and the National Institute of Geographic and Forest Information (IGN).

- Regions are the apex level territorial division of France, which is presently divided into 22 such zones [49].
- Departments are administrative units of France governed by an elected body, i.e., the departmental council. There are currently 96 departments in France.
- Arrondissements are the next level of subdivision of the departments and organize the local police, fire department, and occasionally elections. There are currently 332 arrondissements in France.
- Communes are the smallest and the oldest French administrative unit [50], administered by the municipal council and headed by a mayor. Currently, there are about 36,000 communes in mainland France.
- Urban Units are formed on the basis of contiguous built-up areas [51], and typically merge neighboring communes denotes by urban continuity.
- IRIS is a fine-grained territorial subdivision of France, such that the number of inhabitants in each IRIS zone is around 2,000. This definition is employed by the French National Institute of Statistics and Economic Studies (INSEE) for statistical analyses [52, 53].

Social-Economic indicators. INSEE collects data on population, education, income and other socio-economic statistics on a quinquennial basis across the whole France. The institute has an open-data website that includes, among others, the information below.

- Population contains data for different age groups for both men and women from different years at Commune [54], and IRIS [55] levels.
- Average income encompasses median and quantiles of consumption units per household at commune [54] and IRIS levels for various years. In addition, the data also include Gini indexes that provide information on income inequality between individuals and households in a given region.
- Educational level information includes the number of people attending school or university, in communes [54] and IRIS zones [56], for different age groups.

Telecommunications. The National Agency for Radio Frequencies (ANFR) and the Authority for the Regulation of Electronic Communications (Arcep) are French regulatory bodies in the telecommunication area, and gather data such as coverage and antenna location for electronic communications. Some potentially relevant datasets are made available by the agencies, as listed below.

- Operator coverage is a cartographic platform that assembles all geographic data related to different mobile networks (2G, 3G, 4G, 5G) for all operators [57].
- Radio and antennas location data is handled by ANFR, which updates its mobile network deployment observatory monthly and lists all radio sites authorized on French territory via a cartographic platform [58].

6 CONCLUDING REMARKS

This paper presents a dataset of mobile traffic made available to the research community within the context of a challenge, which features for the first time service-level demands, as well as unprecedented spatial resolution and geographical coverage in a developed country. We believe that the NetMob23 dataset will inspire researchers to design and implement a number of innovative analyses, and discover new knowledge about how and why people consume mobile applications.

ACKNOWLEDGEMENTS

The mobile network traffic and coverage data presented in Section 2.2 and Section 2.3 were collected during the research project CANCAN (Content and Context based Adaptation in Mobile Networks), grant no. ANR-18-CE25-0011, funded by the French National Research Agency (ANR). The NetMob 2023 Data Challenge is organized with the support of the research project NetSense, grant no. 2019-T1/TIC-16037 funded by Comunidad de Madrid, and of the research project CoCo5G (Traffic Collection, Contextual Analysis, Data-driven Optimization for 5G), grant no. ANR-22-CE25-0016, funded by the French National Research Agency (ANR).

REFERENCES

The NetMob23 Dataset

https://doi.org/10.1145/3366423.3380141


