Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 231 (2024) 161–167

Procedia Computer Science

www.elsevier.com/locate/procedia

The 14th International Conference on Emerging Ubiquitous Systems and Pervasive Networks (EUSPN 2023)
November 7-9, 2023, Almaty, Kazakhstan

From Raw Data to Informed Decisions: The Development of an Online Data Repository and Visualization Dashboard for Transportation Data

Ioannis Tsouros[1*], Amalia Polydoropoulou[1], Athena Tsirimpa[1], Ioannis Karakikes[1], Shahram Tahmasseby[2], Anas Ahmad Nemer Mohammad[3], Wael K. M. Alhajyaseen[2]

1 University of the Aegean, Department of Shipping, Trade and Transport, 82132, Chios, Greece

2 Qatar Transportation and Traffic Safety Center, Department of Civil and Environmental Engineering, Qatar University, P.O.Box 2713, Doha, Qatar

3 Department of Civil and Environmental Engineering, Qatar University, P.O.Box 2713, Doha, Qatar

Abstract

This paper presents the design, implementation, and practical use of a specialized online data repository and an interactive visualization of a transportation dashboard. Specifically tailored to handle and interpret large-scale transportation data within the Qatari context, the combined platform serves as a comprehensive solution for managing extensive datasets, including GPS traces from taxis and e-scooters, which are examined as primary use-cases in this paper. The online data repository provides a centralized hub for efficient data storage and management of public transport including Mobility-as-a-Service (MaaS) data. Concurrently, the visualization dashboard fosters an intuitive, user-friendly interface for data exploration and analysis. Through real-world applications within Qatar's transportation ecosystem, we elucidate how these innovative developments can inform data-driven policy decisions in crucial areas such as infrastructure development, resource allocation, and safety measures. Ultimately, this study underscores the pivotal role of effective data management and advanced visualization in maximizing the potential of big data, providing valuable insights for urban mobility planning and enhancing the landscape of policy-making in Qatar and worldwide.

Keywords: Transportation Data Repository; Urban Mobility; Visualization Dashboard; Big-data; Transport policy-making

* Corresponding author. Tel.: +0-000-000-0000 ; fax: +0-000-000-0000 .
E-mail address: jtsouros@aegean.gr

1877-0509 © 2024 The Authors. Published by Elsevier B.V.
This is an open access article under the CC BY-NC-ND license (https://creativecommons.org/licenses/by-nc-nd/4.0)
Peer-review under responsibility of the Conference Program Chairs
10.1016/j.procs.2023.12.188

1. Introduction

Transportation has emerged as a suitable and fertile field for the application of big data analytics due to the massive amount of data generated by various transport services, vehicles, and users. The ability to collect, manage, and analyze this data can significantly enhance our understanding of mobility patterns, inform transportation policy, improve transport system efficiency, and ultimately contribute to more sustainable and inclusive cities.

Despite the potential benefits, managing large-scale transportation data presents a multitude of challenges, ranging from data storage and management to data security, privacy, and ethical issues. Moreover, the value of data is not in its raw form but in the insights that can be extracted from it.

A common pitfall is the tendency to accumulate data without having a clear strategy for analysis or understanding of what insights might be beneficial. Thus, it is the process of converting this raw data into actionable information that can inform decision-making that is the pivotal step in the data management process. This calls for meticulous data processing, as well as the application of rigorous data analytics methodologies.

In addition, big data can often be perceived as an 'absolute truth', but this is not necessarily the case. Issues such as data quality, accuracy, and completeness can all affect the insights derived. The potential for errors or biases in the data, whether they arise due to technical issues during data collection.

This paper presents the development of an Online Data Repository and Visualization Dashboard specifically designed for transportation data, more specifically for the Qatar QNRF MaaS project. The Online Data Repository provides a centralized, secure, and scalable solution for storing and managing large volumes of transportation data, while the Visualization Dashboard serves as a flexible and easy-to-use tool for exploring, analyzing, and presenting this data in an intuitive and meaningful way.

Our work aims to demonstrate practical and scalable steps toward bridging the gap between raw transportation data and policymaking by providing a comprehensive tool that would enable efficient data management, powerful data analysis, and effective data communication. This dual approach may warrant that transportation data is not just collected and stored, but also processed, visualized, and translated into insights that can inform and influence transportation policy and planning. In the following sections, we will detail the processes, challenges, and solutions encountered in the development of the Online Data Repository and Visualization Dashboard, providing a blueprint for similar endeavors in the field of transportation. The intention is to stimulate further discourse and development in the area of data-driven decision making, contributing to the advancement of sustainable, resilient and efficient transportation systems.

2. Literature review

In the transportation sector, among others, the effective management and analysis of data play crucial roles in system performance and policy-making [1]. However, handling large amounts of data often demands robust data repositories, well-documented data processes, and adaptable and flexible modules of data processing. A well-managed data repository not only stores and organizes data, but also maintains data consistency, security, and privacy [2]. Within the sector, data repositories have been utilized for various purposes including traffic management, routing and scheduling, and policy planning [3]. However, the development of data repositories that can effectively manage the vast and diverse nature of transportation data is still a challenging task, calling for solutions that can deal with issues related to data volume, velocity, variety, and veracity [4].

At the same time, turning raw data into actionable insights involves several processes including data cleaning, processing, analysis, and visualization. Among these, visualization plays a particularly crucial role as it represents data in a more tangible and understandable manner, enabling decision-makers to recognize patterns, trends, and anomalies [5]. In the transportation context, dashboards and other visualization tools have been increasingly recognized for their utility in presenting complex mobility patterns and trends in intuitive ways, aiding in both real-time decision-making and long-term planning [6].

Mobile applications have become an indispensable tool in the collection of travel data due to their widespread use and the rich dataset they can provide. These applications, which can be installed on smartphones, can passively collect geolocation data in real time, providing valuable insights into travel patterns, routes, and timings [7,8] They can also actively engage users in providing further data, such as mode of transport, purpose of trip, and travel experience [8]

The integration of data repositories and visualization dashboards presents an innovative solution to the challenges of handling transportation data [9]. An integrated approach provides a seamless transition from data collection to interpretation, reducing the risk of data loss or misinterpretation during transfer between systems [10]. Moreover, the combined system provides a user-friendly interface that enables even non-technical users to extract useful insights from the data [11]. While the advantages of an integrated data repository and visualization dashboard are many, challenges remain particularly regarding data privacy and security [12]. Transportation data often include sensitive information, such as individual location traces and travel diaries, that need to be carefully managed to avoid potential breaches of privacy or misuse of data [13]. Therefore, the implementation of robust security measures and privacy-preserving techniques is an essential aspect of developing these systems [14].

Big data technologies have shown significant promise in improving the transportation sector, helping make sense of the vast amounts of data generated by vehicles, infrastructure, and users [15]. However, while the opportunities presented by big data are substantial, it is essential to be aware of its limitations and challenges. As noted by several scholars, big data is not a solution to all problems, which is why its use should be guided by clear, specific objectives, and supported by appropriate methodological and technical capabilities [16, 17].

3. Requirements Elicitation and Analysis

The development of both the online data repository and the visualization dashboard began with a comprehensive process of identifying and documenting its requirements. These can be broadly categorized into:
Functional Requirements: These define the core tasks and operations the system should perform. For the repository, it includes tasks like data import/export, storage, and user account management. The dashboard focuses on data visualization, filtering, and user interaction. Non-Functional Requirements: These relate to the system's performance and behavior, emphasizing scalability, reliability, usability, performance, and other quality attributes.

Table 1 Example of elicitation process of functional and non-functional requirements

Requirement Type	Requirement Description	Elicitation Method	Source (Stakeholder/Document)
Functional	Ability to import data from various sources	Interview	Project Manager
Functional	Ability to export data in various formats	Interview	Data Analyst
Functional	Ability to store and manage large amounts of data	Questionnaire	Technical Team
Functional	Ability to search and retrieve data based on various criteria	Observation	End-User
Functional	Ability to visualize data using charts, graphs, maps, etc.	Workshop	All Stakeholders
Non-Functional	Scalability: Handle an increasing amount of data and users	Interview	Technical Team
Non-Functional	Performance: Quick data access and retrieval	Questionnaire	End-User
Non-Functional	Usability: User-friendly interface that is intuitive and easy to use	Observation	End-User
Non-Functional	Reliability: Stable and dependable service with minimal downtime	Workshop	All

The elicitation process involved collaborative meetings with stakeholders, notably the Qatar Ministry of Transport (MoT) delegates. Their feedback was invaluable, providing insights into the end-users' needs and shaping the repository and dashboard's design.

Following this, an iterative evaluation was conducted, comparing the system's capabilities against the documented requirements. Adjustments were made based on feedback, ensuring the final tools effectively met end-users' needs. The evaluation incorporated insights from field experts, further strengthening the system's utility and efficiency.

In essence, this rigorous process of requirements identification, stakeholder collaboration, and iterative evaluation played a pivotal role in crafting tools that are robust, user-centric, and adept at managing and visualizing complex transportation data.

Online Data Repository

The development of the online data repository is rooted in its essential design, tailored to meet the extensive data needs of the transportation sector. Key aspects of its design include:

- **Hardware and Software**: A robust infrastructure supports large data volumes, with specific server specifications, storage capacities, and computational power. The choice of PostgreSQL as the database management system stems from its scalability, security, and open-source nature, making it ideal for this project.
- **Data Management**: The repository facilitates data import from diverse sources, especially large datasets like GPS traces. Emphasis is placed on ensuring data integrity through preprocessing steps such as cleaning and normalization. Efficient data organization allows for quick searches and retrievals.
- **Infrastructure Design**: The repository's design emphasizes robustness and scalability. It employs a UAEGEAN server equipped for large-scale data analytics tasks, focusing on quick data retrieval and efficient system performance.
- **Security and Integrity**: PostgreSQL's features like ACID compliance guarantee reliable database transactions, ensuring data integrity. Its robust security measures offer protection against unauthorized access, further enhanced by our rigorous authentication protocols and user action logs.

The subsequent step involved importing data from various sources, with an emphasis on large datasets such as the GPS traces from taxis and e-scooters. Given the extensive size and complexity of these datasets, the data import process presents several challenges. These include issues with data format compatibility, data integrity, and the handling of incomplete or missing data. To overcome these challenges, we implemented a series of data preprocessing steps, including data cleaning, normalization, and validation, to ensure the accuracy and consistency of the imported data. Finally, we organized the data within the database in a way that facilitates efficient search and retrieval. This involved creating appropriate indexes on the tables and designing optimized query strategies. We implemented an intuitive categorization of data based on attributes such as date, time, location, and vehicle type, to ensure efficient search and retrieval operations.

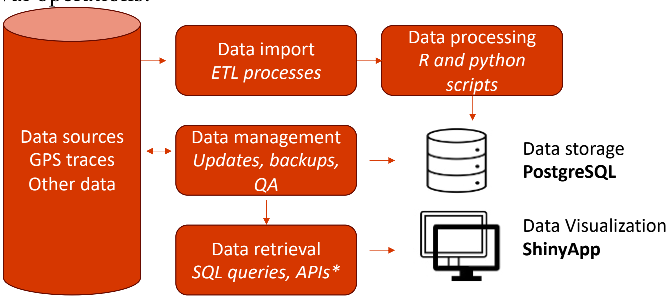

Figure 1. Data flow in the repository

Figure 1 exhibits a visual representation of the data flow within the online data repository. It demonstrates the process from initial data collection, through data import and processing, to eventual data storage and management in a PostgreSQL database. The diagram further illustrates the use of SQL queries and APIs (still under development and open as docks for future usage) for data retrieval and how the retrieved data is visualized using ShinyApp in R.

Our system is designed as a 'centralized, secure, and scalable solution'. The centralized nature is evident from the unified data repository that aggregates data from multiple sources, as depicted in Figure 1. Its scalable architecture, shown in Figure 1, ensures adaptability to growing data volumes. Importantly, we prioritize security: our system employs robust authentication mechanisms, encryption at rest and in transit, and rigorous access controls to protect sensitive data. Managing access rights and user roles in a complex system such as an online data repository is indeed a crucial aspect of its operation and security. When considering access rights, it's necessary to identify the different categories of users who will be interacting with the system and what their interaction entails. These categories may include:

1. **Administrators**: These users shall have the highest level of access and control over the system. They will have the authority to create, modify, and delete records within the repository, manage user accounts, assign roles, and adjust access rights. The administrators will also be responsible for the maintenance and integrity of the system, dealing with technical issues, and ensuring data security.
2. **Government Officials**: This user group includes policymakers and other government personnel who would be using the repository to draw insights from the data. Their access might be limited to retrieving and visualizing data, with no rights/permissions to alter or delete records. They may also have restrictions on which datasets they can access based on their role and responsibilities. This level of access should be carefully controlled to prevent unauthorized data manipulation.
3. **Data Providers**: This category represents those who supply the data to be stored in the repository, such as transportation service providers. These users should have the ability to upload new data or update existing data that they have provided. However, they typically should not be able to alter or delete data provided by others or access sensitive or confidential data.
4. **Read-only Users**: Certain users might only need to view or download the data without making any changes. This access level is often given to users who need the data for analysis or reporting purposes but do not need to alter the data in any way.

In our online data repository, we have implemented strict authentication protocols and regular audits to guarantee the appropriate use of access rights. To ensure a secure environment, each user must authenticate their identity using a secure login procedure that involves multifactor authentication. This process provides an additional layer of security, reducing the risk of unauthorized access.

Moreover, our system is designed to track and log all user actions, thereby fostering a culture of accountability. Any activity within the repository – e.g. data retrieval, modification, or deletion - is recorded along with the user details and timestamp. These logs serve as an essential tool for potential review and audit, enabling us to monitor data access and usage closely.

4. Building the Visualization Dashboard using ShinyApps in R

Using Shiny [18], a package in R, we developed an interactive web application for data visualization hosted on ShinyApps. The dashboard was designed based on elicited requirements, focusing on data visualization, filtering, and metrics representation.

- **GPS Data**: Interactive maps for tracking individual vehicles and heatmaps to represent vehicle density.
- **Temporal Data**: Line graphs and bar charts to identify patterns and trends over time.
- **KPIs and Metrics**: Pie charts, bar graphs, and scatter plots to depict variables like total distance traveled and

average speed.

- **Filtered Data**: Dynamic visualizations updated based on selected criteria, such as time of day or vehicle type.

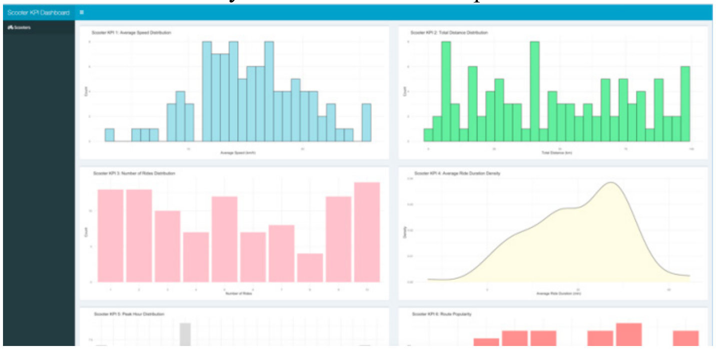

Figure 2. KPIs menu.

The inclusion of these specific visualizations ensures that the dashboard can effectively represent the rich datasets available in the project, facilitating a comprehensive understanding of the data and aiding in informed decision-making. The dynamic nature of these visualizations, facilitated by the ShinyApps platform, also empowers an interactive user experience, promoting user engagement and enabling tailored data exploration.

The functionality of the KPIs in the dashboard is designed with flexibility and user-centricity in mind. The visualizations of these KPIs are straightforward and intuitive, enabling even non-technical users to quickly understand the presented information.

One of the strengths of this dashboard is its ability to be customized and extended. Should users propose new KPIs, these can be incorporated into the dashboard with relative ease, thanks to the modular design of the backend. In fact, this open-ended design specification invites continuous evolution and enhancement of the tool.

The navigation through the KPIs has been designed to be as user-friendly as possible. By simply selecting the desired category from the menu (e.g., "Taxis", "Scooters"), users can view a list of relevant KPIs. Each KPI is presented in its own visual element, allowing users to focus on one metric at a time without distraction. This design facilitates efficient navigation and enables users to easily compare KPIs and understand their interrelationships, making the dashboard an indispensable tool for transportation analysis and decision-making.

An important aspect of the dashboard development was its integration with the online data repository. This required the establishment of a connection to the PostgreSQL database to retrieve the required data for visualization. The integration would activate real-time visualization, enabling the dashboard to reflect the current state of the data repository.

In conclusion, the use of ShinyApps in R for the development of the visualization dashboard provides a robust and flexible tool capable of fulfilling the complex data visualization requirements of the project. The iterative process of development, integration, deployment, and evaluation resulted in a user-friendly dashboard that efficiently transforms raw data into visually appealing and understandable formats, thereby aiding decision-making processes.

Supporting Policy Making and Decisions through Data Visualization

The transition to data-driven decision-making is becoming increasingly essential in policy-making processes. The ability to visualize and understand large and complex datasets would pave the path for more informed decisions and provide empirical evidence to support policy changes. The developments presented in this paper, specifically the online data repository and visualization dashboard, are designed to aid this transition by providing a platform for comprehensive data analysis and visualization.

To highlight the practical implications of our developments, we will explore a few examples utilizing the datasets discussed in this paper and present a couple of user stories:

Improving Transportation Infrastructure: By Using GPS data from taxis and e-scooters, policymakers can analyze the routes frequently taken by these vehicles. High-density zones on the heatmaps may indicate popular destinations or areas of heavy traffic. This information can be utilized to plan infrastructure improvements, such as determining where to construct new roads, bike lanes, or scooter parking areas.

Optimizing Resource Allocation: The visualization dashboard's ability to display vehicle availability at different times can aid in optimizing resource allocation. For example, if data shows a consistently high demand for e-scooters during peak hours in certain areas, policymakers can decide to increase the number of available e-scooters during these times, ensuring a better balance of supply and demand.

Planning Public Transportation Schedules: The temporal data visualizations can also assist in planning public transportation schedules. For instance, if data reveals that certain areas have a high demand for taxis in the early morning, policymakers may decide to adjust operating bus or tram services during this period to meet this demand and reduce congestion.

Enhancing Safety Measures: Analyzing data i.e. average speed and total distance traveled can help identify potential safety concerns. If certain areas consistently show high average speeds (more than permitted speed or generally high speeds such as >120km/h), it could indicate a need for speed limit enforcement or additional safety measures such as speed bumps or traffic-calming infrastructure.

By leveraging the capabilities of the online data repository and the visualization dashboard, policymakers can gain valuable insights from transportation data. This supports the creation of data-driven policies that better address the needs of the community and lead to more efficient and sustainable transportation systems.
In the following section, we'll delve into two user stories that illustrate the potential usage of the dashboard. One of a transportation analyst exploring daily and weekly transport rhythms, and another of a policymaker, whose goals are centered on infrastructure improvements. Both narratives underscore the usefulness and the practical implications of the system.

User Story: A Transportation Analyst Examining Mobility Patterns

- As a transportation analyst working for the Ministry of Transport, I want to understand daily and weekly mobility patterns of our public transport system.
- When I log in to the online data repository and dashboard, I can select the 'Taxis' or 'Scooters' categories from the menu.
- I then navigate to the 'Ride Frequency' KPI to observe the number of rides for each day of the week.
- I also explore the 'Peak Hour' KPI to identify the busiest hours of the day.
- I can export these data for further analysis in my preferred statistical software.
- As a result, I gain insights into mobility patterns which can inform decisions regarding scheduling and resource allocation in our public transport system.

2. **User Story: A Policy-Maker Looking to Improve Infrastructure**

- As a policymaker, I am interested in identifying areas of the city that may require additional transportation infrastructure.
- I log in to the dashboard and select the 'Taxis' or 'Scooters' categories to observe the patterns specific to these transport modes.
- I navigate to the 'Route Popularity' KPI to see which routes are predominantly taken by these motorists.
- Using the map visualizations, I can pinpoint areas with a high concentration of vehicular usage.
- As a result, I can make evidence-based decisions on where to invest in infrastructure improvements, such as dedicated scooter lanes or taxi stands, to better serve the public.

5. Conclusions

The primary objective of this study was to create an online data repository and a visualization dashboard to assist in the analysis and interpretation of large-scale transportation data. The data repository serves as a centralized storage system capable of handling extensive data volumes, including big datasets such as GPS traces from taxis and e-scooters. In parallel, the visualization dashboard presents an interactive tool that empowers policymakers and non-technical individuals to extract valuable insights from complex datasets. The integration of the repository and the dashboard fosters an efficient data flow, enabling real-time visualization and decision-making.
The potential impact of the developments detailed in this paper on policy-making is substantial. The ability to visualize large and complex datasets aids policy-makers in making data-driven decisions, resulting in policies that are better aligned with actual needs and circumstances. As demonstrated through the real-world examples, insights derived from the datasets can guide improvements in transportation infrastructure, resource allocation, transportation schedules, and safety measures. This ultimately leads to the enhancement of urban mobility and the overall quality of life for residents.
Transportation service providers also stand to take advantage of the advancements presented in this study. With access to the online data repository and visualization dashboard, service providers can better understand the travel behavior diaries and spatial/temporal patterns of their vehicles. For instance, information about routes frequently taken, peak demand times, and preferred vehicle types can all be extracted from the system. Such insights can guide service modifications, resource allocation, and even pricing strategies, thereby improving service efficiency and customer satisfaction. Additionally, service providers can leverage this system for internal management and operational purposes. The dashboard can assist in identifying potential operational issues, such as the clustering of vehicles in certain areas or unusually high speeds, enabling service providers to resolve them.
Despite the comprehensive approach and findings of this study, some limitations should be noted. The data repository primarily focuses on taxis and e-scooters, suggesting that including a broader range of transportation modes, such as buses or bicycle-sharing programs, might offer a richer understanding of urban mobility patterns. Furthermore, cultural factors, as demonstrated in the e-scooter usage analysis within the Qatari environment, can greatly influence transportation patterns, and these nuances might not be entirely captured solely through data. Additionally, the efficiency of the dashboard depends on the user's proficiency, implying that non-technical individuals might need further training to maximize its potential. While the developments presented in this paper have proven to be effective in managing and visualizing transportation data, the nature of technology and data science suggests that there will always be room for further improvement and innovation. Future work could focus on incorporating additional datasets, such as bus location data, into the repository and refining the visualization techniques as new methods emerge. Additionally, further user

feedback should be solicited to continuously improve the usability and functionality of the dashboard.
From a managerial perspective, the tools and findings of this study carry substantial significance. Insights on mobility patterns provide managers with the knowledge to ensure efficient resource allocation, catering to areas of higher demand and optimizing service delivery. They can also adjust operational strategies based on real-time data. For example, during high-demand periods or in zones with peak usage, introducing dynamic pricing could be beneficial. By identifying unusual patterns or high-speed areas, managers can prioritize and implement safety measures, enhancing public safety and potentially influencing public perceptions positively. Furthermore, an understanding of user behaviors and preferences empowers managers to initiate targeted marketing or loyalty campaigns, fostering better customer relationships. Lastly, data on frequently traveled routes or congestion areas can guide infrastructure investment decisions, such as the creation of dedicated scooter lanes or taxi stands, ensuring an optimal user experience.
In conclusion, this study emphasizes the importance of effective data management and visualization in leveraging the potential of big data for transportation planning and policy-making. The developed online transportation data repository and visualization dashboard provide a robust toolset for understanding complex datasets of transportation, paving the way for informed, data-driven decision-making in the realm of urban mobility.

Acknowledgment:
This report was made possible by an NPRP award [NPRP 13S-0130-200211] from the Qatar National Research Fund (a member of the Qatar Foundation). The statements made herein are solely the responsibility of the authors.

References

1. Cavoukian, A., & Chibba, M. (2018). Start with privacy by design in all big data applications. Guide to big data applications, 29-48.
2. Kitchin, R. (2014). The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences. Sage.
3. [3] Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. International Journal of Information Management, 35(2), 137-144.
4. Batty, M. (2013). Big data, smart cities and city planning. Dialogues in Human Geography, 3(3), 274-279.
5. Few, S. (2009). Now You See It: Simple Visualization Techniques for Quantitative Analysis. Analytics Press.
6. Chen, M., Mao, S., & Liu, Y. (2014). Big data: A survey. Mobile Networks and Applications, 19(2), 171-209.
7. Montini, L., Prost, S., Schrammel, J., Rieser-Schüssler, N., & Axhausen, K. W. (2015). Comparison of travel diaries generated from smartphone data and dedicated GPS devices. *Transportation Research Procedia*, *11*, 227-241.
8. Chirico, F., Fermi, F., Fiorello, D., Martino, A., Tsirimpa, A., Kamargiannis, K., ... & Tsouros, I. SMARTPHONE-BASED TRAVEL SURVEY IN TURIN: A NEW DATA-DRIVEN APPROACH TO ASSESS URBAN MOBILITY PATTERNS.
9. Liu, R., & Li, Y. (2018). Exploring the relationship between urbanization and energy consumption in China using ARDL (autoregressive distributed lag) and FDM (factor decomposition model). Energy, 150, 527-543.
10. Sun, Z., Du, Y., Wang, Y., & Fu, Y. (2016). Visualization in big data: A survey. Neurocomputing, 214, 778-791.
11. Kitchin, R., & McArdle, G. (2016). What makes big data, big data? Exploring the ontological characteristics of 26 datasets. Big Data & Society, 3(1), 2053951716631130.
12. Mayer-Schönberger, V., & Cukier, K. (2013). Big data: A revolution that will transform how we live, work, and think. Houghton Mifflin Harcourt.
13. Cavoukian, A., & Castro, D. (2014). Big Data and Innovation, Setting the Record Straight: De-Identification Does Work. Toronto, ON: Ryerson University.
14. Boyd, D., & Crawford, K. (2012). Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. Information, communication & society, 15(5), 662-679.
15. Chen, M., Mao, S., & Liu, Y. (2014). Big data: A survey. Mobile Networks and Applications, 19(2), 171-209.
16. [Kitchin, R. (2014). Big Data, new epistemologies and paradigm shifts. Big Data & Society, 1(1), 2053951714528481.
17. Kitchin, R., & McArdle, G. (2016). What makes big data, big data? Exploring the ontological characteristics of 26 datasets. Big Data & Society, 3(1), 205395171663113
18. Chang, W., Cheng, J., Allaire, J.J., Sievert, C., Schloerke, B., Xie, Y., Allen, J., McPherson, J., Dipert, A., & Borges, B. (2023). Shiny: Web Application Framework for R. Retrieved from https://shiny.posit.co/; https://github.com/rstudio/shiny.
19. ggplot2: Wickham, H. (2016). ggplot2: Elegant graphics for data analysis. Springer-Verlag New York. https://ggplot2.tidyverse.org