

UNIVERSITÀ DEGLI STUDI DI MILANO-BICOCCA

DATA SCIENCE LAB IN SMART CITIES
FINAL ESSAY

Comparative Analysis of Open Data Ecosystems
A Case Study of the London DataStore and Comune di Milano

Authors:

Antonio Mastroianni - 898723 - a.mastroianni5@campus.unimib.it
Pasquale Formicola - 893791 - p.formicola@campus.unimib.it

January 30, 2025



Abstract

This report compares the open data ecosystems of the London DataStore and the Comune di Milano Open Data Portal, focusing on their development, adherence to open data principles, and practical utility for urban decision-making. The analysis shows that while both platforms offer a significant amount of publicly available data, the London DataStore provides a more polished user experience, greater data detail, and built-in analytical tools that are accessible to the general public. On the other hand, Milan's portal stands out for its availability of geo-referenced datasets but falls short in providing certain essential datasets and ready-to-use insights. These shortcomings raise questions about whether the missing information is intentionally withheld or if they stem from challenges in data collection and digitization. The findings indicate that Milan's open data strategy could benefit from improved accessibility, enhanced data detail, and the integration of analytical tools to boost public engagement and policy effectiveness.

1 Introduction

The concept of "smart cities" has gained significant attraction in recent years as a revolutionary approach to urban development, aiming to enhance the quality of life for citizens through the integration of technology, innovation, and data-driven solutions. One of the key elements of smart city development is open data. It plays a vital role in promoting transparency, encouraging innovation, and engaging citizens. Open data refers to the free and accessible sharing of information collected by public institutions, which can be used, reused, and distributed by anyone without restriction.

Despite its transformative potential, not all open data is created equally. The effectiveness of open data initiatives depends on several key attributes that determine its quality, usability, and overall impact. Poorly structured, outdated, or inaccessible datasets can hinder innovation and frustrate users, undermining the benefits that open data promises to deliver. This highlights the need for a deeper understanding of what constitutes "good" open data and how cities can effectively implement and manage open data ecosystems to support their smart city goals.

This study examines the principles and characteristics of high-quality open data by exploring established theoretical frameworks and technical standards. To ground these insights in practice, the research aims at comparing the cities of Milan and London on their open data ecosystems. These cities were selected due to the contrasting positions of their respective countries in the Global Open Data Index [1] which provides a unique opportunity to investigate what in the two open data ecosystems may lack compared to other and to identify areas for improvement.

By assessing the policies, technical implementations, and usability of open data platforms in London and Milan, this study aims to identify actionable strategies for evaluating open data ecosystems. The findings will contribute to a deeper understanding of how cities can leverage open data to foster innovation, enhance transparency, and improve urban living.

2 Understanding Open Data in Smart Cities

"Open data refers to data presented in open formats that individuals can use freely and share for any purpose" [2]. To fully harness its potential in smart cities, it is essential to understand the open data value chain[3], a framework that outlines the stages through which data progresses to create impact.

As shown in Fig.1, the process begins with data production, where public institutions, private organizations, or individuals gather and generate data. At this stage, it is crucial to ensure quality, accuracy, and completeness to facilitate effective use later on. After data is produced, the next step involves converting raw data into open formats, focusing on accessibility, interoperability, and licensing. This transformation makes the data usable and widely available to the public, promoting inclusivity and allowing various stakeholders to access the information. The third phase highlights the use and analysis of data, where researchers, developers, and policymakers interact with the information to tackle specific challenges or seize opportunities. This stage includes activities like data visualization, modeling, and integrating data into decision-making processes, turning raw data into actionable insights. Ultimately, the value chain leads to generating impact. The insights gained from open data are utilized to create tangible real-world results, such as improved governance, better public services, innovation-driven economic growth, and increased citizen engagement. By adhering to this structured approach, cities can enhance the flow of data and ensure that open data systems provide measurable benefits to urban environments. The domains in which open data can have impact on a smart city are several [4]. For example, in the context of urban development, open data serves as a catalyst for innovation, transparency, and citizen engagement. By making data accessible, cities empower developers, researchers, and policymakers to create solutions that address urban challenges, enhance public services, and improve the quality of life for residents. Moreover,

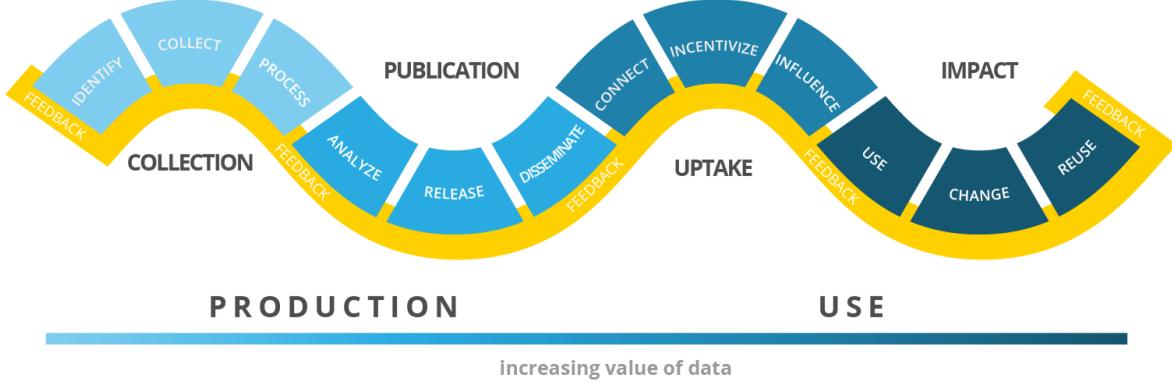


Figure 1: Open Data Value Chain. Source: Open Data Watch

open data significantly enhances urban development by promoting transparency and accountability. When governments openly share data, citizens can monitor public sector activities, leading to increased trust and civic participation[5]. From the economic perspective, open data serves as a catalyst for growth: entrepreneurs and businesses can utilize this freely available information to develop innovative products and services, thereby stimulating economic development and creating job opportunities [6]. In terms of decision-making, open data provides policymakers with evidence-based insights. Access to comprehensive and up-to-date data enables more effective urban planning and resource allocation, resulting in efficient and responsive governance [7].

2.1 Characteristics of Quality Open Data

For open data to reach its full potential in smart cities, it needs to meet certain characteristics that guarantee its quality, accessibility, and usability. The Open Data Charter[8] highlights six essential principles that characterize "good" open data. These principles offer a framework for evaluating the effectiveness of open data systems and act as a standard for cities aiming to improve their open data ecosystems.

1. **Open By Default:** Data should be publicly available unless there are legitimate privacy or security concerns.
2. **Timely and Comprehensive:** Data must be current and complete to ensure relevance and usefulness.
3. **Accessible and Usable:** Data should be easy to access, machine-readable, and accompanied by clear documentation to promote inclusivity.
4. **Comparable and Interoperable:** Data must adhere to common standards to enable integration and cross-analysis across datasets.
5. **For Improved Governance and Citizen Engagement:** Open data should enhance decision-making, transparency, and civic participation.
6. **For Inclusive Development and Innovation:** Data should support equitable growth, empower diverse stakeholders, and drive innovation.

These six principles together outline what makes open data "good", ensuring it is not just accessible but also useful, impactful, and fair. In the field of smart cities, following these principles is essential for harnessing the transformative power of open data to improve urban life and promote sustainable development. For these reasons, these will be the guiding principles of our work. However, many of these principles are inherently challenging to assess directly, particularly when evaluating the practical implementation of open data portals. Factors like citizen engagement, economic impact, and the real-world applications of open data are complex to measure in a standardized and quantifiable manner as this kind of information is not often spread. Thus, the focus of this work will be on exploring the various aspects that derive from these principles. By examining factors such as the accessibility and usability of datasets, the motivations behind the creation of the portals, and the tangible impact of the data on urban development, this analysis will provide a comprehensive understanding of how each city's open data ecosystem operates in practice. Through this approach, the work will highlight the strengths and areas for improvement in the open data ecosystems of London and Milan, while considering both the theoretical underpinnings and practical realities of open data in smart cities.

3 Open Data in London and Milan: A Comparative Analysis

In this chapter, we examine and compare the open data portals of London, specifically the London DataStore[9], and the open data website of Milan's municipality[10]. Our goal is to evaluate their effectiveness, adherence to open data principles, and their overall influence on urban innovation and governance. Through this analysis, we aim to identify best practices and areas that need improvement, which can inform future advancements in open data strategies. We will proceed in three stages:

1. The historical context of each portal will be explored to understand the motivations, stakeholders, and objectives that shaped their creation. This step is crucial to highlight how foundational decisions influence the structure and priorities of the ecosystems today.
2. An attempt to evaluate the two ecosystems according to the six principles will be made. This theoretical assessment will evaluate how well the portals conform to recognized standards for open data quality.
3. A sub-set of the datasets provided by the two portals will be analyzed in practical terms. By choosing similar domains and conducting hands-on evaluations, this study will evaluate the quality, usability, and relevance of the data for typical urban challenges. This practical aspect guarantees that the analysis reflects the real-world usefulness of the portals, enhancing the theoretical framework.

This chapter seeks to present a well-rounded analysis by integrating historical, theoretical, and practical viewpoints to evaluate the two open data ecosystems. Although there are challenges in assessing specific principles or acquiring complete datasets, the study strives to provide practical insights that highlight the portals' roles in fostering urban innovation and governance.

3.1 History of the Open Data Ecosystems

In this subchapter, we explore the origins of the London DataStore and Milan's Open Data Portal, focusing on the motivations, key players, and historical contexts that shaped their development. By analyzing these foundational aspects, we aim to understand how initial decisions influenced the structure, priorities, and impact of these open data ecosystems.

3.1.1 London DataStore

The London DataStore was established in 2010 by the Greater London Authority (GLA) to provide a centralized platform for public data access [11]. Its main goal was to enhance transparency and accountability in government while encouraging innovation through the public sharing of city-related data.

This initiative received strong backing from Mayor Boris Johnson, who viewed it as a chance to make government data accessible and beneficial for citizens, businesses, and developers. The project was spearheaded by GLA Chief Executive Leo Boland and researcher Emer Coleman, who played key roles in conceptualizing and implementing the Datastore. The initiative was also shaped by significant international and national policy developments, such as the UK's Power of Information Taskforce which underscored the potential of open data to enhance governance and foster innovation.

The platform made a notable impact by inspiring the creation of numerous applications and services that directly benefited Londoners. For instance, developers utilized data from the Datastore to create apps that assisted residents in navigating public transportation[12][13], monitoring air quality, and finding available housing. The Datastore also empowered citizens and journalists to examine government operations, resulting in greater public engagement and civic participation.

The Datastore also enabled citizens and journalists to scrutinize government activities, leading to increased public involvement and civic engagement. Over time, it evolved into a crucial resource for addressing urban issues, enhancing service delivery, and fostering public confidence in government. Additionally, the success of the London DataStore motivated other cities to implement similar open data initiatives, solidifying its role as a trailblazer in the global open data movement. By encouraging collaboration among public agencies, private developers, and the community, the Datastore established a standard for how cities can leverage open data to promote transparency and innovation.

3.1.2 Comune di Milano Open Data

The Comune di Milano's Open Data Portal project started in 2012[14] as part of the city's commitment to the "Data-Driven Milan" program[15] and officially launched in 2016. This initiative seeks to improve transparency, accountability, and citizen engagement by making municipal data accessible to the public. The program received

strong support from local government leaders, emphasizing the importance of open data in fostering innovation and collaboration among citizens, businesses, and public institutions.

Key figures involved in the portal's implementation include members of the Technological and Digital Innovation Division of the Municipality of Milan. They have been instrumental in crafting data strategies and integrating datasets into a Linked Open Data (LOD) framework. This initiative is in line with wider European policies that advocate for open government data, aiming to boost public trust and foster civic engagement.

Over the years, the Comune di Milano has implemented several advanced data visualization projects aimed at democratizing access to information. Notable projects include Open Bilancio[16], which provides detailed insights into municipal finances, and the Council's Mandate Report that quantitatively assesses the performance of elected officials during their terms.

The success of the Open Data portal has positioned Milan as a national leader in open data initiatives, comparable to major international cities like London and New York. The municipality continues to prioritize enhancing the portal by developing high-quality datasets based on standardized ontologies and engaging with stakeholders to ensure that the data meets community needs.

3.1.3 History Comparison

The London DataStore and the Comune di Milano Open Data Portal backgrounds highlight different foundation motivations, stakeholder interactions, and effects that are shaped by the specific circumstances surrounding their establishment.

London DataStore's primary focus was to enhance transparency and accountability while fostering innovation by providing developers, businesses, and citizens with access to public data. On the other hand, Milan's Open Data Portal aimed to democratize access to municipal data and emphasize civic engagement. Both initiatives were championed by prominent leaders who recognized the transformative potential of open data and relied on collaborations among policymakers, technical experts, and civil society, reflecting a multi-stakeholder approach. Both ecosystems have had significant impacts on their respective cities even if on different aspects. The London DataStore inspired the creation of apps that assist residents with navigating public transport, monitoring air quality, and accessing housing information. On the other hand, Milan's portal has reportedly supported initiatives more orientated towards decision-making.

3.2 Open Data Platforms

In this chapter, the open data portals of London and Milan are evaluated to determine their alignment with the six principles of the Open Data Charter.

By looking at the design, functionality, and content of these portals, we seek to grasp how well they reflect the principles of transparency, accessibility, and innovation that are essential to open data ecosystems. This assessment reveals both the strengths and weaknesses in the approaches of the two cities, offering insights into their roles in urban development, civic engagement, and fostering public trust.

3.2.1 Open By default

This principle determines that all data should be accessible to the public, without having to ask officials for the needed information.

Both the London DataStore and the Comune di Milano Open Data Portal exhibit a commitment to this principle at first glance. Each platform hosts a significant number of datasets covering a wide array of topics and reaching, in total, a similar amount of available datasets. This breadth suggests a foundational adherence to the idea of making government data accessible.

However, a deeper examination reveals notable differences in the granularity and usability of the data provided. For instance, the London DataStore offers highly detailed datasets in critical domains such as crime and transportation. For example, for what concerns crime data, it provides monthly data for each borough[17], allowing for nuanced analyses of trends and localized issues. In contrast, the Comune di Milano Open Data Portal only publishes yearly crime data aggregated at the city level. This limited granularity makes it difficult for citizens, researchers, and policymakers to explore neighborhood-specific patterns or respond to issues in a timely manner. These differences illustrate varying degrees of adherence to the "Open by Default" principle. Although both platforms technically offer transportation data, the limited depth and granularity of Milan's datasets could restrict their practical usability and effectiveness raising questions on which are the reasons for which these kind of information are not available. In contrast, the London DataStore showcases a more comprehensive approach, delivering detailed and actionable data that closely aligns with the objectives of this principle.

The practical implications of these differences will be examined in the next section, where we will analyze specific datasets from both platforms in detail. This analysis will demonstrate how the level of detail and usability of data affect the platforms' ability to foster urban innovation and tackle challenges unique to cities.

3.2.2 Timely and comprehensive

Ensuring data is timely allows stakeholders to make informed decisions, while comprehensiveness ensures that datasets provide a full and accurate representation of the subject matter. Given the broad scope of the datasets hosted by both portals, it is impractical to expect every dataset to meet this principle equally. Some datasets, such as historical archives or static geographic data, naturally do not require frequent updates, while others, like transportation or crime data, demand more regular revisions to remain relevant.

Upon review, there is no clear indication that either platform does not respect the "Timely and Comprehensive" principle in their data offerings. Both portals offer a variety of datasets with different update frequencies, and for many key areas, the data seems to be refreshed according to practical and administrative needs.

3.2.3 Accessible and Usable

Both the London DataStore and the Comune di Milano Open Data Portal serve as centralized data hubs that align with the principle of providing machine-readable datasets. As it is possible to see in Fig.2, they offer data in commonly used formats like CSV, JSON, GeoJSON, XLSX and XML, which ensures that various analytical tools and software can easily allow easy accessibility towards developers, researchers, and civic organizations.

The user experience of accessing and navigating these datasets varies between the two platforms. The London DataStore stands out by providing a dedicated analysis section organized by domain. This feature simplifies the process for users belonging to general public to find datasets that match their interests and also offers pre-processed data along with insights drawn from the raw information. For example, users can explore curated reports and visualizations on important themes such as housing[18], air quality[19] and quality of life[20], making it easier to extract actionable insights from the data.

3.2.4 Comparable and Interoperable

Both portals show a dedication to this principle, yet they take different approaches, each offering distinct strengths and weaknesses.

For the London DataStore, most datasets are organized to present information at the borough level. This level of detail enables users to easily compare data from different areas of the city and allows for the merging of datasets for more in-depth analysis thus enhancing the usability of the data for both citizens and developers.

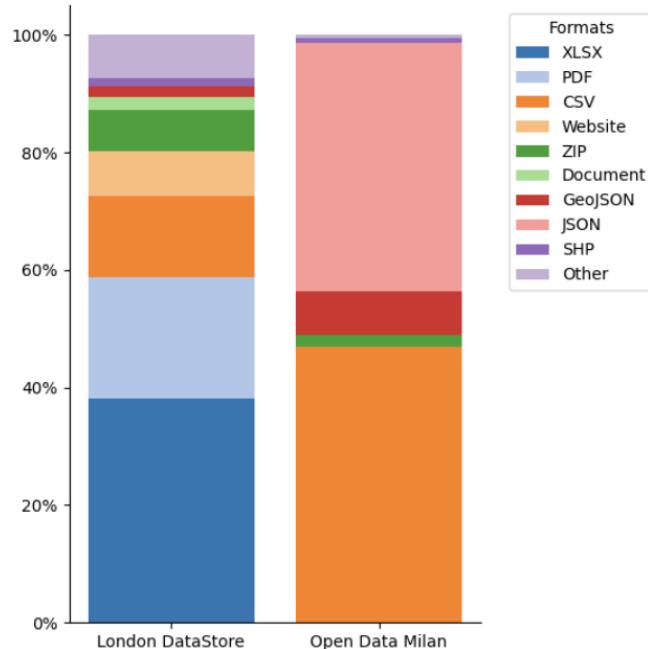


Figure 2: Data format distribution for each portal

In contrast, the Comune di Milano portal, while not always providing data on a neighborhood level, offers a significant amount of georeferenced datasets. This geospatial data allows analysts or developers to perform spatial joins and integrate datasets based on geographic references. For instance, users can merge datasets with a base map or geographic reference dataset to derive insights at a finer spatial scale. This geospatial focus, as will be shown in the next section, enables powerful analyses that leverage location-based data to address urban challenges.

3.2.5 For Improved Governance and Citizen Engagement

As already mentioned, London DataStore makes available a set of analyses and tools directly on its portals. By presenting information in a user-friendly format, the platform makes it easier for the general public to access and understand the data, allowing citizens to engage with it without needing advanced technical skills. This transparency boosts civic engagement and fosters trust in governance by making government activities and urban issues clearer and more actionable.

Its correlative for Milan instead lacks the same depth of analysis and ready-to-use tools for general users. Much of the data on the Milan portal requires additional processing and technical expertise to derive meaningful insights. This limitation may reduce the platform's immediate utility for citizens seeking transparency and quick answers to civic questions.

3.2.6 For Inclusive Development and Innovation

Both portals play a role in economic and social development through open data, yet they vary in terms of accessibility and usability. The London DataStore not only supplies raw data but also provides pre-analyzed insights, making it simpler for policymakers, businesses, and the public to utilize information for innovation. Furthermore, it encourages collaboration by bringing together various organizations and researchers.

On the other hand, Milan's portal, while abundant in geo-referenced datasets beneficial for urban planning, lacks structured reports or analysis tools, which means users must handle data on their own. This can hinder accessibility for non-experts and diminish the platform's immediate effectiveness. Additionally, some vital datasets that could promote inclusive development are either missing or not detailed enough.

3.3 Platforms Data

In this section, we explore a more hands-on evaluation, comparing the datasets of the two open data ecosystems. The goal is to evaluate how useful and valuable the information from both portals is for specific urban Decision-making tasks. These tasks will showcase the advantages and disadvantages of each platform in terms of practical applications.

The comparison will focus on three key areas:

1. **Schooling:** Evaluating how easily data can be accessed and utilized to monitor city-related aspects of education, such as school locations, enrollment statistics, with the aim of identifying opportunities for improving governance and resource allocation.
2. **Crime:** This task examines the availability and detail of crime-related datasets. We will compare how the two portals present crime data, assessing the granularity, frequency, and comprehensiveness of the information. This evaluation is crucial for effective policing, public safety, and transparency.
3. **Environment:** The environmental dimension is increasingly vital for sustainable urban development. This comparison will focus on datasets related to green areas, bike stalls, and electric car charging stations, which contribute to sustainability. By evaluating how well each portal supports access to these datasets, we aim to assess their potential to inform policies promoting sustainability, improve quality of life, and address climate challenges.

By comparing these specific tasks, this chapter aims to offer a practical assessment of the two portals' capacity to support urban development, governance, and innovation through open data. Through this analysis, we can better understand how each portal contributes to solving real-world urban challenges.

3.3.1 Schooling

The first step to assess this task for the City of Milan was to consider the Nuclei di Identità Locale (NIL)[21] as a geographic reference. They were chosen because as they are officially recognized by the Municipality of Milan as they serve as standardized spatial divisions, and they were also available on the respective open data portal.

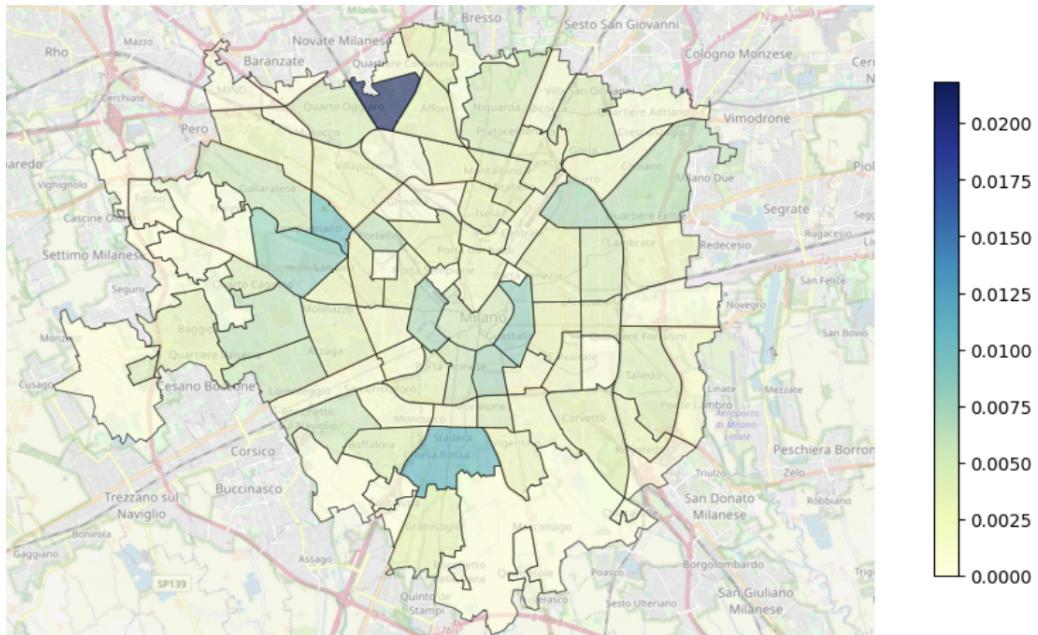


Figure 3: School density per NIL

In the demographic section of the portal, it was also possible to access to a dataset containing the number of residents of each NIL, a key information that was used to calculate some of the indexes that will be later discussed.

The open data portal for Milan provides a wide range of geo-referenced datasets that allow to access information such as schools position and a broad range of information about their students such as their age, gender

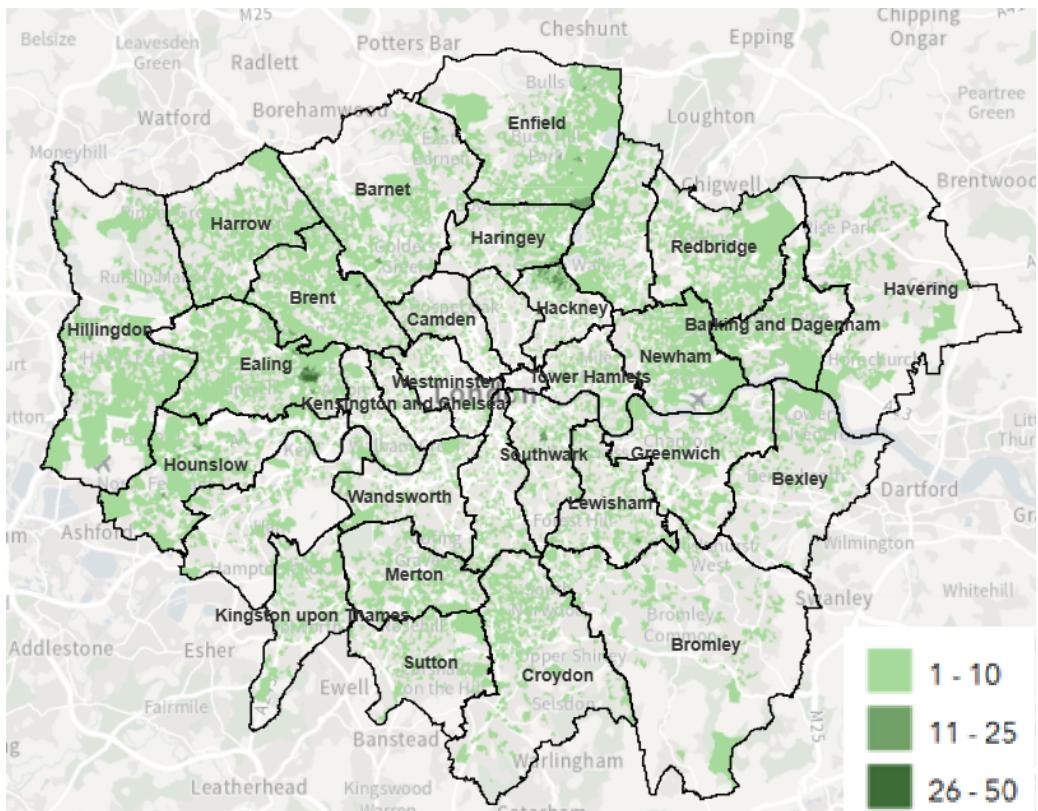


Figure 4: English language deprivation from 3 to 15 years old

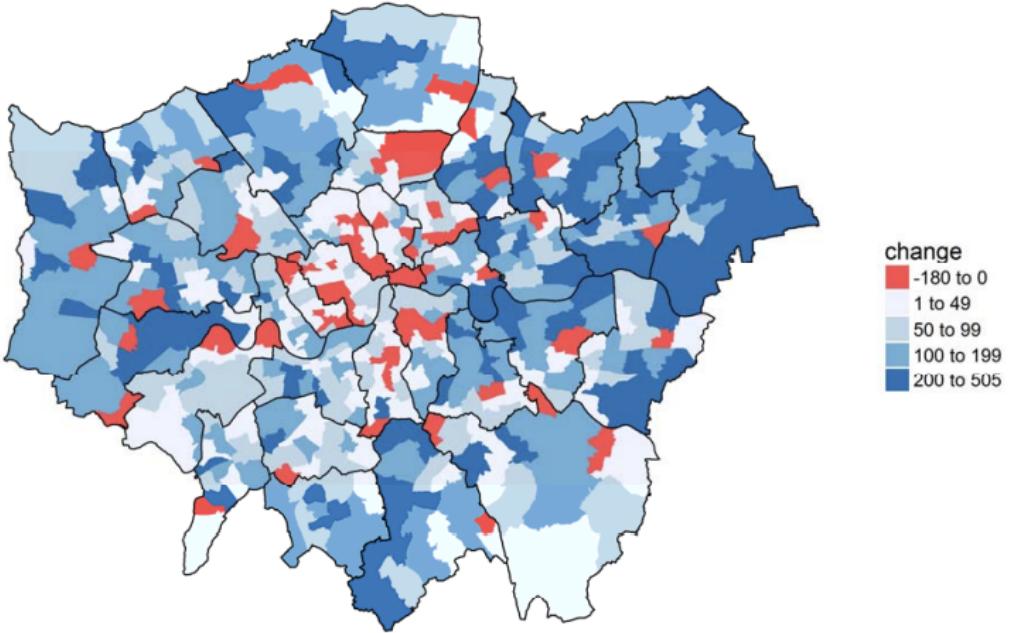


Figure 5: : Change in state place demand 2016/17 to 2027/28 by ward (secondary school) - Source: London DataStore

and nationality. We have figured that a possible task for a municipality officer concerned about the theme of education could be school subscription overcrowding. In that case, a good metric to monitor could be the number of schools per pupil (individuals between 6 and 18 years old) as the following index.

$$\text{School Density Index} = \frac{\text{Number of Schools}}{\text{People between 6 and 18}}$$

As it is possible to see in Fig.3, the map provides valuable insights for urban planning and education management. This metric helps identify areas where schools may be overcrowded, ensuring that resources are allocated effectively to address gaps in educational infrastructure. It also highlights disparities in school accessibility, supporting equitable education opportunities across neighborhoods. Additionally, the index aids in integrating educational needs into broader urban planning efforts and allows policymakers to monitor trends over time, ensuring that interventions remain effective and responsive to population changes.

The London DataStore does not directly provide georeferenced data on school positions, but it offers a wealth of other information such as detailed data about students and deprivation levels. This data enables analyses at a granularity finer than the borough level, the ward, providing a nuanced view of deprivation across the city. Using this data, we have been able to analyze the deprivation levels across London at a granularity finer than the borough level, as shown in Fig.4. This level of detail is crucial for understanding localized inequalities and ensuring that educational resources are directed toward areas where they are most needed, thus promoting equitable access to quality education.

Moreover, the platform offers analyses that are crucial for decision-making. For example, there is a study[22] which predicts shifts in student numbers across various boroughs and school types. These kind of projections, like the one shown in Fig.5, are essential for strategic planning, allowing policymakers to foresee future demand and distribute resources effectively, ensuring that the education system evolves with demographic trends and continues to serve the city's needs efficiently.

3.3.2 Crime Data

The analysis of crime data from open data portals shows notable differences in detail and usability between the two cities. As mentioned earlier, a detailed look at the crime data section on Milan's open data portal indicates that the only dataset available reports the total number of crimes committed in Milan each year divided by type. With this lack of information, the only valuable insight that it is possible to gather might be given just

by performing tasks such as tracking the trend of a specific crime through the years, as in Fig. 6a or retrieving the most committed crime during a specific year as shown in Fig.6b

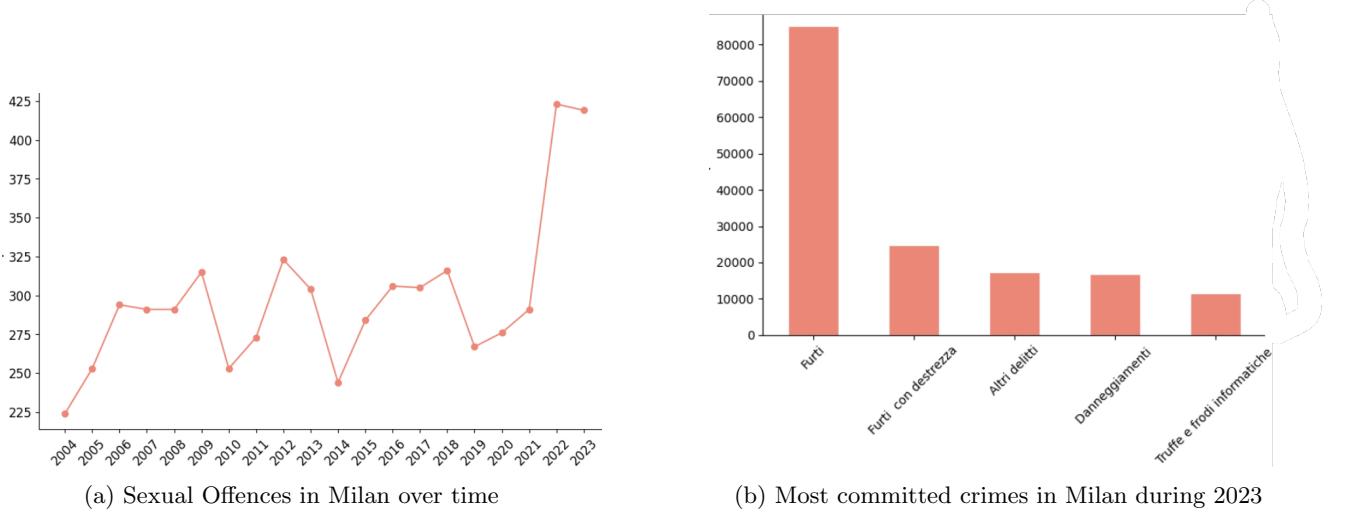


Figure 6: Possible crime insights derived by Milano open data about crime

In contrast, the London DataStore offers much more detailed crime data. The platform provides specific information, such as the type of crime, the month it occurred, and the exact borough where it took place. This level of detail allows for the extraction of more useful and accurate insights, making the data more effective for policy-making and raising public awareness. For instance, as illustrated in Fig. 7, it was possible to pinpoint the neighborhoods with the highest number of sexual offenses in 2024. This capability underscores the value of comprehensive crime data in facilitating targeted interventions, resource distribution, and community safety efforts.

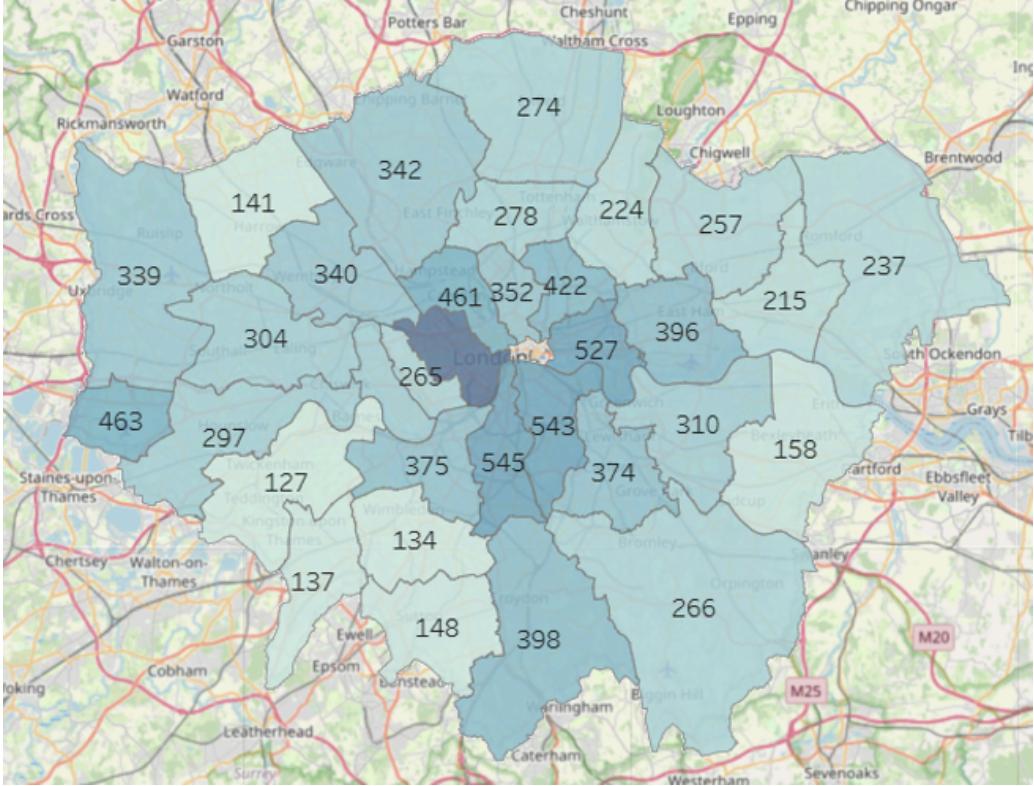


Figure 7: Sexual Offences in London per borough in 2024.

In addition, The Greater London Authority, has created with this and other source data a secure platform known as SafeStats[23], which compiles a wide array of crime and community safety datasets from various

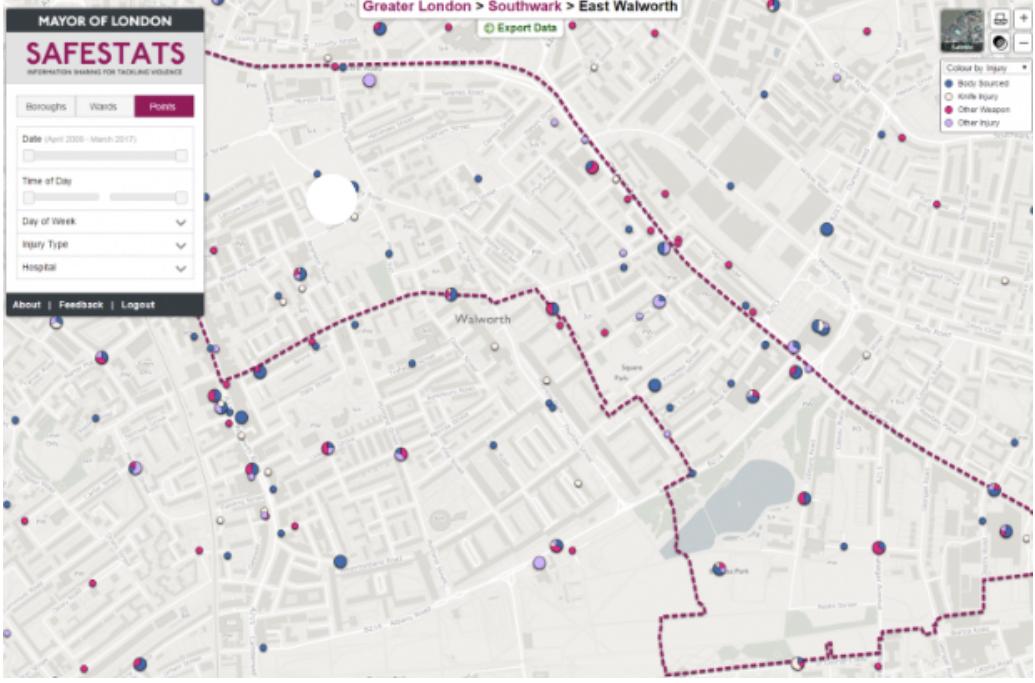
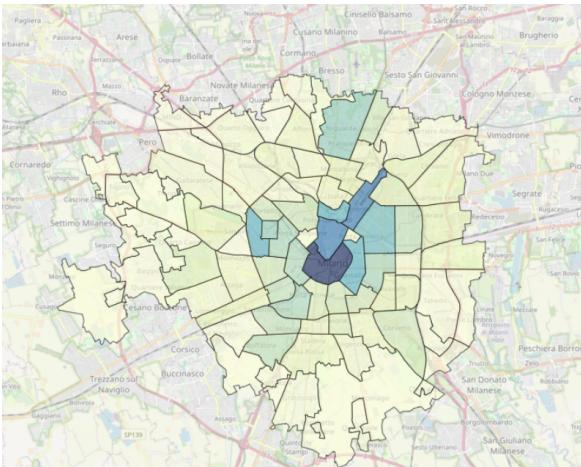


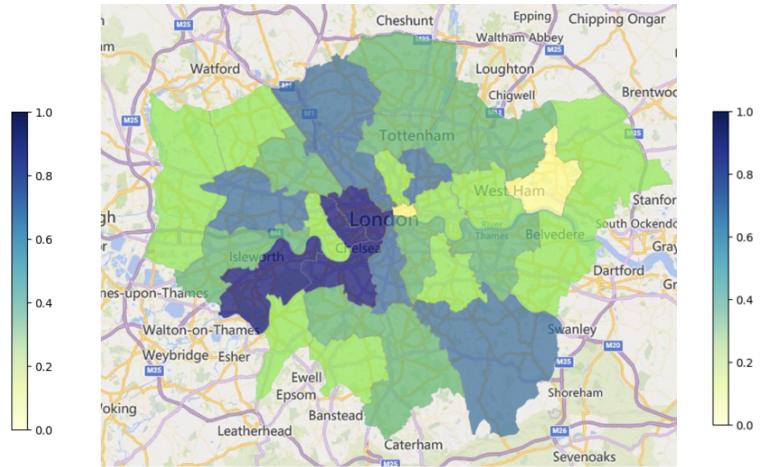
Figure 8: Crime locations visualized with SafeStats - Source: London DataStore .

organizations, including the Metropolitan Police Service and the London Ambulance Service. SafeStats contains over 18 million data points from the last decade, allowing to visually see crime locations on the city map as shown in Fig.8 and enabling thorough strategic analysis of crime trends and community safety challenges. Its main goal is to improve public safety by providing detailed insights into crime patterns and supporting initiatives like the Violence Reduction Unit, which takes a public health approach to address violent crime.

While SafeStats is only accessible to authorized personnel within public safety agencies for security reasons, its development highlights London's comprehensive strategy to utilize data for better governance and public safety. The platform exemplifies effective data sharing and analysis, showcasing how open data can promote transparency and facilitate targeted interventions. This strategy illustrates the sophisticated data infrastructure that supports London's crime data ecosystem.



(a) Sustainability index in Milan for each NIL



(b) Sustainability index in London for each borough

Figure 9: Sustainability index

3.3.3 Environment Data

For Milan, we were able to merge several datasets related to sustainability, which provided valuable information on various aspects of the city's environmental infrastructure. These datasets included data on green areas, the number of charging stations, and bike stalls, all organized per NIL. By combining these datasets, we calculated a sustainability index, which, as shown in Fig.9a can serve as a useful tool for urban planners and decision-makers to evaluate and monitor the environmental performance of different areas within the city.

$$\text{Sustainability Index} = 0.333 \times \text{Green Area per km}^2 + 0.333 \times \text{Bike Stalls} + 0.333 \times \text{Charging Stations}$$

Such index can be particularly useful for urban development and policy-making, as it helps to identify areas that might need further investment in green spaces, cycling infrastructure, or electric vehicle charging stations. As shown in Fig.9b, the same index was calculated exactly in the same way for London, demonstrating the same availability of data for the two platforms.

4 Results

In this section, we present the results of our comparative analysis between the London DataStore and the Comune di Milano Open Data Portal under the three aspects that were chosen for the comparison.

4.1 Ecosystems History

The London DataStore and the Comune di Milano Open Data Portal were both established to enhance transparency, improve governance, and encourage innovation through public data sharing. However, their backgrounds and the reasons for their creation differ in several important ways.

The London open data platform aimed to make city-related data publicly accessible to promote transparency, foster innovation, and empower citizens, businesses, and developers. Its development was significantly influenced by national and international policies. The platform quickly gained popularity and became a model for other cities around the globe, leading to various applications that enhanced urban services like transportation, housing, and public safety.

On the other hand, the Comune di Milano Open Data Portal represents the city's commitment to open government and has played a crucial role in promoting transparency and engaging citizens. While Milan's platform shares similar objectives with London's, it lacks the same historical significance or widespread influence. Nevertheless, it has made notable progress in providing accessible datasets on urban services, particularly focusing on geographic data and sustainability.

Despite the differences in their historical contexts, both platforms share a common goal: to leverage open data to enhance governance and tackle urban challenges. However, London's initiative is more advanced in scale and reach.

4.2 Adherence to the Open Data Principles

Both the London DataStore and the Comune di Milano Open Data Portal are dedicated to making urban data accessible to the public, but they approach this goal in different ways. While both platforms generally follow the core principles of open data, there are significant differences in how they implement these principles, particularly regarding the depth and usability of the data.

At first glance, both platforms seem to embrace the idea of being open by default, offering a wide range of datasets on several topics concerning smart cities. However, a closer look shows that the London DataStore provides more detailed and granular datasets, which enhances its overall usefulness. For instance, London DataStore crime data is available at a finer level and is more frequently, allowing for more accurate tracking of crime trends over time. In contrast, Milan's portal only offers annual city-wide crime data, which limits the ability to analyze trends with the same level of detail. This difference points to a potential shortcoming in Milan's capacity to provide real-time, localized insights for effective decision-making.

In terms of comparability and interoperability, both platforms perform admirably. They provide data in machine-readable formats, making it possible to use the datasets in various applications. However, London stands out in its ability to compare data across different boroughs, thanks to the consistent level of detail and standardization of data points. Milan, while it offers many georeferenced datasets, encounters difficulties in facilitating direct comparisons across neighborhoods, as the data is not always uniformly structured or detailed at a smaller scale.

Another area where London's platform excels is in its emphasis on usability. The London DataStore goes beyond merely providing raw data; it also offers a range of analysis tools and visualizations that help users extract actionable insights.

4.3 Datasets

The Comune di Milano Open Data Portal offers a substantial amount of geo-referenced data. This georeferencing allows for a high level of spatial analysis, making it possible to examine various datasets in conjunction with one another, thus providing a more holistic understanding of urban trends. However, the real added value in Milan's platform lies primarily in the availability of raw data that requires further processing and analysis by users. In many cases, users need to combine multiple datasets to derive meaningful insights, which could be more time-consuming for those without technical expertise.

In contrast, the London DataStore enhances user experience by offering accessible, pre-processed data and insights directly on its portal. It covers essential topics like education, crime, and sustainability, presenting analysis, visualizations, and projections that the general public can easily utilize. These insights are designed to be user-friendly, effectively bridging the gap between raw data and practical knowledge. For example, the London DataStore features predictive tools and analysis reports, such as those forecasting future school place demands or crime trends, which are available to anyone with internet access, making the platform more user-friendly for both policymakers and citizens.

While Milan offers valuable geospatial data that can be analyzed using external tools, London's platform allows for a faster and more straightforward application of data in decision-making processes. This distinction emphasizes a significant difference: Milan's portal is better suited for users with technical skills who can manipulate and integrate the data themselves, while London's platform provides a more comprehensive and user-friendly experience, making its insights easily accessible to a wider audience.

In summary, although both portals present strong data offerings, the London DataStore excels in ease of use, granting users immediate access to meaningful insights without requiring extensive effort to process the data. Conversely, while Milan's data is rich and detailed, it often necessitates additional steps for users to derive actionable insights, which can hinder its accessibility and immediate value.

5 Conclusions & Future Works

This analysis has revealed several important differences between the open data ecosystems of the Comune di Milano and the London DataStore. While both platforms aim to provide public access to city-related data, the London DataStore stands out for its superior range, accessibility, and user-friendliness of datasets. London's platform not only features a wide variety of data but also offers processed insights, making it easier for the public and policymakers to draw actionable conclusions. In contrast, the Comune di Milano's platform, often requires more technical expertise and external tools for effective analysis. Furthermore, many insights that are easily accessible on London's platform are not presented in a similar manner for Milan.

A major issue that came up during the evaluation is the lack of essential data from the Milan portal. Important datasets, like in the case of crime data, either do not provide the same level of detail or granularity found in the London DataStore or are simply unavailable. This raises the question: why are these datasets not accessible? Could it be that this information is being withheld, or is there a more significant problem with data collection, management, or digitization in Milan? This situation is still unclear and needs further investigation.

One possibility is that the city of Milan might encounter difficulties in obtaining or digitizing specific datasets. An unorganized and inefficient data collection process could impede the city's capacity to deliver data promptly and thoroughly. Additionally, there could be political or strategic motives behind the decision to withhold certain types of data, although such choices should ideally be clear and justifiable to the public.

In any case, it is clear that Milan's open data portal is some steps behind the London DataStore. To bridge the gap with the London platform, a clear commitment to transparency should be reinforced by ensuring that high-value datasets are publicly available. Currently, crucial data seem to be missing or restricted, raising questions about whether this is due to policy decisions, technical limitations, or institutional reluctance. Addressing these gaps would be a major step forward in democratizing access to information. Moreover, Milan could enhance user engagement by integrating analytical tools directly into the portal. London DataStore does not just provide raw data; it also offers processed insights and interactive reports that make complex information more digestible for policymakers and citizens alike. By following this model, Milan could encourage data-driven governance and public participation. Finally, fostering collaboration between public institutions, private entities, and the research community could drive further improvements. London DataStore has thrived by working with various stakeholders to refine its datasets and expand its applications. Milan, as there are no evidences this is already

happening, could develop similar partnerships to ensure continuous updates, better integration of mobility and environmental data, and the creation of predictive models for urban planning.

Future work could focus on exploring these potential barriers in more depth—investigating whether the absence of certain data is due to technical, political, or strategic reasons. Additionally, further comparisons could be made with other leading open data cities to identify best practices that could help Milan accelerate its progress in this area. By adopting a more proactive approach to data openness, Milan could unlock the full potential of open data to drive urban development and sustainability in the years to come.

References

- [1] “Tracking the state of open government data,” *Global Open Data Index*, 2024. [Online]. Available: <http://index.okfn.org/place.html>
- [2] P. de Bas, “Indicators for an open data impact assessment,” *data.europa.eu*, p. 8, 2024. [Online]. Available: <https://data.europa.eu/sites/default/files/report/Indicators%20for%20an%20open%20data%20impact%20assessment.pdf>
- [3] OpenDataWatch, “The data value chain: Moving from production to impact,” *Open Data Watch*, 2020. [Online]. Available: <https://opendatawatch.com/publications/the-data-value-chain-moving-from-production-to-impact/>
- [4] A. Ojo, E. Curry, and F. A. Zeleti, “A tale of open data innovations in five smart cities,” *WordPress*, 2015.
- [5] M. Edmiston, S. Coker, S. Jamilla, and T. Tshabalala, “The pros and cons of open data,” *Merl*, 2021. [Online]. Available: <https://merlcenter.org/guides/pros-and-cons-of-open-data>
- [6] J. Glasco, “Open data and innovation: Vast opportunities, untapped value,” *Bee Smart City*, 2023. [Online]. Available: <https://www.beesmart.city/en/smart-city-blog/open-data-and-innovation>
- [7] F. T. Neves, M. de Castro Neto, and M. Aparicio, “The impacts of open data initiatives on smart cities: A framework for evaluation and monitoring,” *ResearchGate*, 2020.
- [8] “Odc principles,” *Open Data Charter*, 2015. [Online]. Available: <https://opendatacharter.org/principles/>
- [9] “London data store,” *Mayor of London*, 2010. [Online]. Available: <https://data.london.gov.uk/>
- [10] “Comune di milano open data,” *Comune di Milano*, 2012. [Online]. Available: <https://dati.comune.milano.it/>
- [11] E. Coleman, “A lesson from the london datastore,” *Beyond Transparency*, 2013. [Online]. Available: <https://beyondtransparency.org/part-1/lessons-from-the-london-datastore/>
- [12] “Transportapi.” [Online]. Available: <https://www.transportapi.com>
- [13] “City mapper.” [Online]. Available: <https://citymapper.com/?lang=en>
- [14] G. Sindoni and M. Massenzio, “An open data platform for decision making in local public administration,” *ResearchGate*, 2023. [Online]. Available: https://www.researchgate.net/publication/372080943_An_Open_Data_Platform_for_decision_making_in_local_public_administration
- [15] O. of the Deputy Mayor for Digital Transformation and C. S. of the City of Milan, “Milano digital transformation,” 2021, documented by the Office of the Deputy Mayor for Digital Transformation and Citizens’ Services of the City of Milan. [Online]. Available: https://www.yesmilano.it/system/files/allegati/paragrafi/16256/MilanDigitalWeek_Factsheets_Mission_032921_v4.pdf
- [16] “Milan open bilancio.” [Online]. Available: <https://analyticsplatform.comune.milano.it/anonymous/extensions/OpenBilancioHome/OpenBilancioHome.html>
- [17] “Local government act 1972,” *Stationery Office Books*, 1972.
- [18] G. Yu, G. Csontos, and J. Gleeson, “Housing in london 2024,” *London Data Store*, 2024. [Online]. Available: <https://data.london.gov.uk/housing/housing-in-london/>
- [19] K. Wilkins, “Air quality fact sheet: Pollutant trends in the uk and london,” *Air Quality Consultant*, 2024.
- [20] P. Toyne, A. Sinha, M. Adebowal-Schwarte, A. Coote, D. Elliott, J. Hirigoyen, S. Heath, M. Mehra, C. Webbe, C. Wood, L. Secretariat, J. Colombneau, K. Guida, and V. Avery, “London’s quality of life indicators report 2017: Measuring london’s progress towards becoming a sustainable world city,” *London Sustainable Development Commision*, 2017. [Online]. Available: https://londondatastore-upload.s3.amazonaws.com/QoL_2017_Report_online.pdf

- [21] “Nuclei di identita’ locale (NIL),” *Comune di Milano*, 2012. [Online]. Available: https://www.pgt.comune.milano.it/sites/default/files/allegati/NIL_Intro.pdf
- [22] GLA, “2018 gla school place demand projections,” *GLAIntelligence*, 2018.
- [23] “Geocoding istv data,” *Mayor of London*, 2016.