CARMA 2026

8th Int. Conf. on Advanced Research Methods and Analytics
1-3 July 2026, Valencia, Spain
DOI: https://doi.org/10.4995/CARMA2026.2026.21988

Accuracy and Representation of Public Real Estate Assets: A Comparative Analysis of Italian Metropolitan Cities

Ida Ciarlillo, Giorgio Ialenti, Giuliano Resce
Department of Economics, University of Molise, Italy.

How to cite: Ciarlillo, I., Ialenti, G., Resce, G. (2026). Accuracy and Representation of Public Real Estate Assets: A Comparative Analysis of Italian Metropolitan Cities. In: 8th International Conference on Advanced Research Methods and Analytics (CARMA 2026). Valencia, 26-28 June 2026. https://doi.org/10.4995/CARMA2026.2026.21988

Abstract

This study examines the alignment between the Ministry of Economy and Finance (MEF) database and the Real Estate Disposal and Enhancement Plans (PAVI) published by the 14 Italian Metropolitan Cities, using 2022 data. Through web scraping we collected official information on more than 17,000 real estate units and integrated it with the MEF database through a multi-key record linkage procedure, supported by a stratified manual validation. The approach updated more than 80% of previously missing entries and reclassified about 7.13% of disposal-use records inconsistently labelled as "Directly Used". The findings show that digital methodologies can substantially strengthen already well-established open-data systems and offer a replicable workflow to improve the quality of official statistics on public real estate.

Keywords: *Public real estate assets; Italian Metropolitan Cities; web scraping; record linkage; open data.*

1. Introduction

Italian public real estate assets constitute a strategic resource whose management and enhancement has been central to the country's political and administrative agenda for two decades. The ability to census, classify, and accurately communicate such assets is a fundamental prerequisite for the rationalisation of public expenditure, the generation of extraordinary revenues, and the revitalisation of the local real estate market.

The Ministry of Economy and Finance (MEF), through its Treasury Department, has published annually until 2022 the Public Administrations' Asset Account, a database containing the inventory of properties owned or used by Italian public administrations. It represents the principal official national-level source of information on public real estate assets.

Article in Press

This work is licensed under a Creative Commons License CC BY-NC-SA 4.0
Editorial Universitat Politècnica de València

In parallel, Italian legislation requires Municipalities to prepare and publish the Real Estate Disposal and Enhancement Plans (PAVI), planning instruments through which local administrations identify properties to be sold or enhanced, with the aim of improving public asset management efficiency and contributing to the rehabilitation of municipal budgets.

Given the existing discrepancies between the information communicated by Municipalities through PAVIs and that reported in the MEF database, we use the self-reporting of local authorities to explore possible improvements. By regulatory definition, PAVI-listed assets are not used for institutional purposes, making their classification as "Directly Used" in the MEF database not fully consistent with their inclusion in a disposal or enhancement plan. The results reveal substantial informational discrepancies in the representation of Italian public real estate assets and point to concrete opportunities for harmonisation.

2. Background

The issue of information quality in the public sector is well documented. Heeks (2002) identifies the "information gap" as a primary cause of e-government policy failure, underscoring how inaccurate or outdated data undermines effective administrative decision-making. In the specific context of public assets, Kaganova and McKellar (2006) highlight the necessity of integrated, accurate, and timely information systems for efficient public real estate management. In Italy, the Court of Auditors (2018; 2020) has repeatedly noted informational deficiencies in the public real estate detection system, attributing them to both organisational factors (staff shortages) and systemic factors (lack of common standards); similarly, the OECD (2015) has identified public real estate governance as an area of potential efficiency improvement.

The role of digital methods and machine learning in improving public-sector data quality and local government management has been increasingly recognised. Antulov-Fantulin, Lagravinese, and Resce (2021) demonstrate the applicability of machine learning approaches to predict financial distress in local governments; Bucci, Ferrara, and Resce (2024) show that efficiency gaps in Italian municipalities are strongly related to fiscal decentralisation patterns; and Resce (2022) underscores the influence of both political and managerial factors on the financial management of local governments. On administrative transparency, Ruijer et al. (2017) describe the paradox of "opaque transparency", where data are formally available yet practically inaccessible or non-comparable, a characterisation that applies directly to PAVIs. Janssen et al. (2012) and Ubaldi (2013) argue that open and structured data formats are a prerequisite for improving public asset management and for enabling advanced analytical processes.

Article in Press

3. Regulatory Framework

3.1. The MEF Database and the Asset Account

The Public Administrations' Asset Account is prepared pursuant to Legislative Decree No. 91/2011 and Presidential Decree No. 132/2013, which regulate the economic-patrimonial accounting of public administrations. The MEF database collects information transmitted annually by thousands of public entities and constitutes the most complete information base available at the national level. Among the surveyed variables is "Asset Use", which indicates whether a property is: (i) directly used by the owner administration; (ii) not used; (iii) unusable; or (iv) under renovation or maintenance. Empty cells indicate the failure, or omission, of an entity to provide such information to the MEF. This variable is crucial for programming public asset disposal and enhancement policies.

3.2. Disposal and Enhancement Plans

PAVIs are grounded in Law No. 135/2012 (the "Spending Review"), which in Article 3 obliges Municipalities with a population exceeding 10,000 inhabitants to prepare and approve by 30 November of each year a real estate disposal and enhancement plan, to be attached to the budget forecast. Legislative Decree No. 85/2010 (Federalismo Demaniale) further expanded the possibilities of transferring state properties to territorial entities, making PAVIs an even more relevant instrument in local asset policy.

A critical shortcoming is the absence of standardisation requirements for PAVI preparation and publication. While Municipalities are obliged to publish, no provisions define: (a) the data format (PDF, spreadsheet, structured dataset); (b) the minimum mandatory information fields; or (c) the designated section of the institutional website. This regulatory gap produces a highly heterogeneous landscape: PAVIs may be found in sections labelled "Budget", "Transparency", "Assets", "Council Resolutions", or in areas that are difficult to identify, with formats and levels of detail varying profoundly across Municipalities.

4. Data and Methodology

4.1. Data Sources

The analysis draws on two main sources: (1) the MEF 2022 database, obtained from the Treasury Department of the Ministry of Economy and Finance, containing the complete inventory of Italian public administration properties with reference to 2022 (over 472,000 records); and (2) the PAVIs of the 14 Italian Metropolitan Cities, collected through systematic research on the institutional websites using web scraping techniques (Bottai et al., 2025).

Article in Press

4.2. Web Scraping Pipeline

The collection of PAVIs was automated through a web-scraping pipeline organised in four sequential stages. (i) Seed identification and crawling: for each Metropolitan City, the institutional domain was used as seed, and a breadth-first crawler explored pages up to a configurable depth, restricting navigation to the official domain and to a curated list of section keywords (e.g. patrimonio, alienazioni, valorizzazioni, bilancio, trasparenza, delibere). A polite crawling policy was adopted (User-Agent identification, request throttling, respect of robots.txt). (ii) Candidate document selection: links were retained when both the anchor text and the URL matched a curated lexicon of PAVI-related terms (e.g. "piano alienazioni", "piano valorizzazioni") and a year-reference filter compatible with the 2022 fiscal exercise; PDF and spreadsheet attachments referenced in Council resolutions were also collected as fallback sources when no thematic page was available.

(iii) Parsing: spreadsheet files were read with structured-data libraries preserving column types, while PDFs were processed through a layout-aware table-extraction routine with fallback to text-based extraction; each extracted record was normalised into a common schema covering address, asset type, intended action (sale, enhancement, lease), and any reported use-classification field. (iv) Error handling: the pipeline implemented automatic retries on transient HTTP errors, structured logging of failed downloads, and per-document validation rules (non-empty key fields, plausible address format, recognised asset-type vocabulary); documents failing parsing were flagged for manual inspection rather than silently discarded, ensuring full auditability.

4.3. Record Linkage Procedure

Properties listed in PAVIs were matched with MEF database records through a multi-key record linkage procedure based on: (i) the ISTAT municipality code; (ii) the address and municipality of location; (iii) the type of real estate asset; and (iv) the name of the holder administration. Given the limited standardisation of PAVIs, the procedure required address normalisation (expansion of abbreviations, removal of punctuation and toponymic prefixes, lower-casing, stripping of civic-number suffixes) and terminological reconciliation of use-type categories through a controlled vocabulary built from the MEF nomenclature.

Matching was carried out in two stages. A deterministic stage attempted exact matches on the full key (municipality code + normalised address + asset type). When the deterministic stage failed, a probabilistic stage was applied: candidate pairs were ranked by a string-similarity score on the normalised address (token-set ratio), combined with categorical agreement on municipality code and asset type. Pairs above a high-confidence similarity threshold were accepted automatically; pairs in a lower-confidence band were routed to manual adjudication; pairs below the lower threshold were rejected. For each matched property, the "Asset Use" value

Article in Press

in the MEF database was extracted. The expected values for PAVI-listed properties were "Not Used", "Unusable", or "Under Renovation/Maintenance". Any other classification, particularly "Directly Used" or an empty cell, was treated as an informational discrepancy.

4.4. Validation and Treatment of Ambiguous Cases

To assess the reliability of the matching procedure, a stratified random sample of matched pairs was drawn, with strata defined by Metropolitan City and by the matching stage (deterministic vs. probabilistic) to guarantee adequate coverage of cities with smaller PAVI volumes and of the more uncertain probabilistic matches. Each sampled pair was independently inspected by two reviewers who verified the consistency of address, municipality, asset type, and holder administration; disagreements were resolved by discussion, and inter-rater agreement was monitored across the sample.

Three categories of ambiguous cases were handled explicitly. Partial-address matches (e.g. street name without civic number, or generic localities such as "various properties") were accepted only when the remaining keys (municipality code, asset type, holder administration) uniquely identified the record on the MEF side; otherwise the PAVI record was reported as "unmatched". One-to-many matches, in which a single PAVI entry plausibly mapped to multiple MEF records (typical of buildings composed of multiple cadastral units), were resolved by grouping at the building level and propagating the PAVI classification to all linked units, with a flag indicating aggregated linkage. Conflicting use classifications across linked records were retained as observations of internal MEF inconsistency and reported separately. All unmatched and flagged records were preserved with an explicit status field, so that subsequent analyses could distinguish genuine misclassification from residual uncertainty in the linkage.

5. Results

Table 1 presents the percentage distributions of MEF "Asset Use" classifications among PAVI-listed properties, facilitating comparison across Metropolitan Cities of very different dimensions.

Table 1. Percentage distribution of properties by MEF "Asset Use" category (2022). Source: authors' elaboration on MEF 2022 data and Metropolitan Cities' PAVIs.

Metropolitan City	% Under Renov./Maint.	% Unusable	% Not Used	% Directly Used	% Empty Cells
Bari	0.00%	0.00%	0.00%	100.00%	0.00%
Bologna	1.94%	2.91%	12.14%	13.11%	69.90%
Cagliari	0.00%	0.00%	4.00%	96.00%	0.00%
Catania	0.00%	0.00%	33.07%	16.85%	50.08%
Florence	0.00%	0.00%	75.00%	25.00%	0.00%

Article in Press

Genoa	0.00%	0.00%	66.67%	11.11%	22.22%
Messina	0.00%	0.00%	100.00%	0.00%	0.00%
Milan	0.00%	0.00%	5.23%	77.78%	16.99%
Naples	0.00%	0.00%	11.63%	0.12%	88.25%
Palermo	0.00%	11.81%	12.20%	42.91%	33.07%
Reggio Calabria	0.00%	0.00%	2.94%	96.73%	0.33%
Rome	0.00%	0.00%	3.44%	54.69%	41.88%
Turin	0.00%	1.03%	14.43%	76.98%	7.56%
Venice	0.00%	3.60%	0.00%	69.06%	27.34%
TOTAL	**0.02%**	**0.26%**	**12.18%**	**7.14%**	**80.40%**

The percentage analysis identifies three distinct informational profiles. A first group, Bari (100.00%), Cagliari (96.00%), Reggio Calabria (96.73%), Turin (76.98%), Milan (77.78%), and Venice (69.06%), is characterised by a high share of "Directly Used" properties (above 50%), indicating a systematic updating problem relative to the actual status of PAVI-listed assets. A second group, Naples (88.25% empty cells), Catania (50.08%), Rome (41.88%), and Bologna (69.90%), presents a predominant lack of information for a large share of surveyed assets. Naples alone accounts for 85.99% of total matched properties (14,688 units). A third group, Florence (75.00% "Not Used"), Genoa (66.67%), and Messina (100.00%), shows a more virtuous profile, with majority shares of correctly classified properties, although on samples of limited size.

Taken together, the figures show that the integration of PAVI-derived information allowed the update of more than 80% of previously missing entries (empty cells in the MEF database) and the reclassification of approximately 7.13% of properties incorrectly recorded as "Directly Used" despite their formal inclusion in a disposal or enhancement plan. The remaining matched properties were either already coherently classified ("Not Used", "Unusable", "Under Renovation/Maintenance") or fell into ambiguous categories explicitly flagged during the validation step.

6. Discussion

6.1. Implications for MEF Database Quality

The high incidence of missing information (80.40%) indicates that the MEF data-collection system would benefit from the methodology proposed here, which can help guarantee the completeness required for evidence-based policy decisions. The classification of PAVI-listed properties as "Directly Used", that is, of assets formally destined for disposal or enhancement, constitutes a misalignment attributable to multiple factors: (i) the absence of automatic updating mechanisms between the MEF database and the Municipal Council resolutions that approve PAVIs; (ii) the fragmentation of the public information system, in which different databases do not communicate with each other; and (iii) the lack of incentives and sanctions for

Article in Press

administrations that fail to promptly update transmitted information. These factors interact with the organisational and staffing constraints already documented by the Court of Auditors (2018; 2020), suggesting that purely technological interventions are unlikely to be sufficient without complementary governance measures.

6.2. PAVI Critical Issues: Absence of Standardisation

The lack of PAVI standardisation produces consequences across three dimensions: retrieval difficulty (the absence of a univocal placement indication makes documents hard to identify, with several PAVIs found exclusively as attachments to Council resolutions); non-comparability (PAVIs range from textual PDF lists to spreadsheet tables with profoundly different detail levels); and poor updatability (most PAVIs are prepared in static format, preventing dynamic updates and integration with other information systems). It should be noted, however, that PAVIs, precisely because they are drafted by institutions directly embedded in the local territory, are potentially more accurate than central databases in reflecting the actual condition of individual assets. Local administrations possess direct knowledge of properties under their jurisdiction, including current use, state of maintenance, and disposal prospects, information that the MEF database can only approximate through aggregated reporting. Rather than treating PAVIs merely as a benchmark against which to measure MEF discrepancies, future policy should invest in elevating them to a primary, machine-readable data source within a unified information architecture.

7. Conclusions

This study has analysed informational discrepancies between the MEF 2022 database and the PAVIs of the 14 Italian Metropolitan Cities. The results show that automated retrieval, parsing, and record linkage techniques can substantially improve the completeness and consistency of public real estate information: more than 80% of previously missing entries were updated, and approximately 7.13% of disposal-use classifications were corrected from "Directly Used" to a status consistent with PAVI inclusion. These findings demonstrate the ability of digital methods to complement, rather than replace, existing administrative data-collection processes, by surfacing inconsistencies that are difficult to detect through manual review alone.

In light of these results, several policy recommendations emerge. First, mandatory standards should be introduced to improve the identification and representativeness of the phenomena, including the adoption of a structured data format and the creation of a dedicated section on institutional websites to ensure transparency and accessibility. Second, verification and control mechanisms should be implemented to ensure the quality and completeness of the data transmitted by local authorities. Finally, the implementation of methodologies such as the one

Article in Press

proposed here fosters the representativeness of information and enables data sources to be updated automatically.

Future research developments may extend the analysis to a multi-year time frame and to the complete universe of Municipalities obliged to publish, in order to verify whether the critical issues identified in Metropolitan Cities are representative of the national picture.

References

Antulov-Fantulin, N., Lagravinese, R., & Resce, G. (2021). Predicting bankruptcy of local government: A machine learning approach. *Journal of Economic Behavior & Organization*, 183, 681–699.

Bottai, C., & Trentini, F. (2025). *Mapping the Italian Third Sector: Insights from Web-augmented Administrative Data*.

Bucci, V., Ferrara, G., & Resce, G. (2024). Local government spending efficiency and fiscal decentralization: evidence from Italian municipalities. *Applied Economics*, 56(5), 599–614.

Court of Auditors. (2018). *Report on the result of control carried out on the management of real estate assets of public administrations*. Rome: Court of Auditors.

Court of Auditors. (2020). *Report on the result of control carried out on the management of real estate assets of public administrations*. Rome: Court of Auditors.

Heeks, R. (2002). Information systems and developing countries: Failure, success and local improvisations. *The Information Society*, 18(2), 101–112.

Janssen, M., Charalabidis, Y., & Zuiderwijk, A. (2012). Benefits, adoption barriers and myths of open data and open government. *Information Systems Management*, 29(4), 258–268.

Kaganova, O., & McKellar, J. (Eds.). (2006). *Managing Government Property Assets: International Experiences*. Washington, DC: Urban Land Institute.

Ministry of Economy and Finance – Treasury Department. (2022). *Public Administrations' Asset Account – Year 2022*. Rome: MEF.

OECD. (2015). *Government at a Glance 2015*. Paris: OECD Publishing.

Resce, G. (2022). The impact of political and non-political officials on the financial management of local governments. *Journal of Policy Modeling*, 44(5), 943–962.

Ruijer, E., Grimmelikhuijsen, S., & Meijer, A. (2017). Open data for democracy: Developing a theoretical framework for open data use. *Government Information Quarterly*, 34(1), 45–52.

Ubaldi, B. (2013). *Open Government Data: Towards Empirical Analysis of Open Government Data Initiatives*. OECD Working Papers on Public Governance, No. 22. Paris: OECD Publishing.

Note. The 14 Italian Metropolitan Cities established by Law No. 56/2014 (Delrio Law) are: Bari, Bologna, Cagliari, Catania, Florence, Genoa, Messina, Milan, Naples, Palermo, Reggio Calabria, Rome Capital, Turin, and Venice.

Funding. The authors gratefully acknowledge the funding provided by "The Enhancement of Municipal Real Estate Assets" under the Agreement with the Ministry of Economy and Finance (project code H33C25000430001).

Article in Press