Crowdsourced Data for Urban Planning: Evaluating OpenStreetMap Categorization Reliability in Historical Urban Centres

Federica Deri(✉), Chiara Anselmi, and Federico Mara

Department of Energy, Systems, Territory and Construction Engineering, University of Pisa, Largo Lucio Lazzarino, 56122 Pisa, Italy
federica.deri@phd.unipi.it

Abstract. The contemporary shift toward proximity-based urban planning relies heavily on high-resolution spatial data. Although Volunteered Geographic Information (VGI) platforms such as OpenStreetMap (OSM) provide extensive and free datasets, the quality of commercial POIs in dense historic centres has not yet been thoroughly examined through a functional taxonomy. This study assesses the spatial and semantic reliability of crowdsourced commercial data by comparing it with ground-truth surveys in the historic centres of Pisa and Lucca (Tuscany, Italy). Using a 23-tier meso-category classification framework, this research examines the degradation of VGI data through quantitative metrics and an analysis of mapping mismatches. The results reveal significant gaps in data coverage and a marked imbalance in Points of Interest (POIs) in both contexts: whilst tourism-oriented and food-related services are systematically over-represented, traditional retail suffers from severe under-mapping. Expanding spatial matching tolerances yields negligible improvements, confirming that the main constraint is a significant time lag rather than GPS inaccuracy. In both cities, the mapping community must contend with the rapid turnover of urban retail: the accumulation of 'digital ghosts' – fuelled by 'excess' and 'replaced/inconsistent' nodes – highlights a systemic failure to eliminate obsolete data. At the same time, a clear spatial-semantic dichotomy emerges: when a physical entity is actually mapped, its semantic categorisation proves to be remarkably accurate, in direct correlation with the ontological clarity of the object. Ultimately, this article suggests that, within these two densely touristic historic centres, OSM currently functions more as a cumulative historical record than as a real-time synchronised digital twin. This finding cautions against the potential structural risks of integrating unverified VGI into advanced urban computational models without accounting for the friction of urban commercial turnover.

Keywords: Volunteered Geographic Information (VGI) · Spatial Reliability · Retail Categories

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2027
O. Gervasi et al. (Eds.): ICCSA 2026 Workshops, LNCS 16765, pp. 377–394, 2027.
https://doi.org/10.1007/978-3-032-30536-7_24

1 Introduction

The contemporary landscape of urban planning relies heavily on computational models and advanced spatial analytics, which require high-resolution data regularly refreshed. As planners and policymakers work to implement sustainable, proximity-based frameworks – most notably the "15-min city" concept [1, 2] – the demand for accurate mapping of essential services, retail nodes, and public amenities has increased. This information is essential for evaluating pedestrian access and simulating how urban environments react to transformative policies [3, 4].

Concurrently, the proliferation of Volunteered Geographic Information (VGI) and crowdsourced platforms, such as OpenStreetMap (OSM), has provided researchers with large amounts of geospatial data at no cost. The consequent assumption that "Big Data" inherently equates to reliable data is deeply flawed and represents a critical issue that urban planners must face: when computational algorithms process unverified crowdsourced data to inform urban policy, the propagation of spatial and categorical errors can lead to misdiagnosed urban deficits and misguided structural interventions [5, 6].

Previous literature has extensively documented the quantitative shortcomings of OSM, primarily focusing on geometric accuracy and coverage in comparison to authoritative datasets [7, 8]. Nevertheless, the dimensions of functional categorization and temporal maintenance of urban amenities remain largely underexplored with significant consequences: frameworks relying on broad functional umbrellas – such as grouping all retail into a generic "Commerce" macro-category [1] – fail to capture the functional diversity and the ontological nuances required for micro-urban analysis [9].

This study bridges this gap by merging rigorous spatial validation – comparing OSM records against exhaustive ground-truth surveys – with a refined functional taxonomy tailored specifically for commercial Point of Interest (POI) records developed by Mara et al. [6]. Focusing on the walled historical centres of Pisa and Lucca (Tuscany, Italy), this paper aims to dissect the specific typologies of errors inherent to VGI in these contexts. Specifically, we investigate the completeness and quality of the OSM database in terms of: (i) volumetric representation; (ii) spatial and semantic reliability; (iii) typologies of mapping mismatches (e.g., surplus, redundant, and replaced nodes); and (iv) categorisation correctness. Our ultimate objective is to critically assess whether crowdsourced mapping, in its current state, possesses the structural, temporal, and categorical integrity necessary to support advanced urban analytics in morphologically complex cities like Pisa and Lucca.

On this vein, the paper is structured as follows: Sect. 2 presents the analyzed case studies; Sect. 3 describes the adopted methodology – ranging from the construction of the two datasets to the processing steps performed for their comparison; Sect. 4 presents and discusses the results obtained; finally, Sect. 5 summarizes the key findings and outlines potential developments for future research.

2 Case Studies: Two Walled Historical Centres

The historical walled centres of Pisa and Lucca (Tuscany, Italy) provided an ideal, bounded laboratory for investigating digital geospatial data reliability, serving as a testing ground that challenges the mapping capabilities of crowdsourced communities. Unlike

sprawling suburban areas where commercial nodes are highly visible and segregated, these medieval urban cores are characterized by dense, mixed-use fabrics that foster social vibrancy, walkability, and commercial vitality in plazas and streets [10].

Divided by the Arno river, Pisa's historical centre acts as a primary commercial and cultural hub characterised by high pedestrian flows driven by tourism and a prominent university population. Lucca is completely encircled by its intact Renaissance walls and presents an extremely compact pedestrian environment where commercial amenities are deeply embedded within narrow street networks (Fig. 1).

Testing OSM's reliability in these areas highlights the vulnerability of crowdsourced data in environments where the rapid turnover of commercial activities, the high density of Points of Interest (POIs), and the frequent repurposing of historic ground-floor spaces create a highly volatile environment for digital mapping.

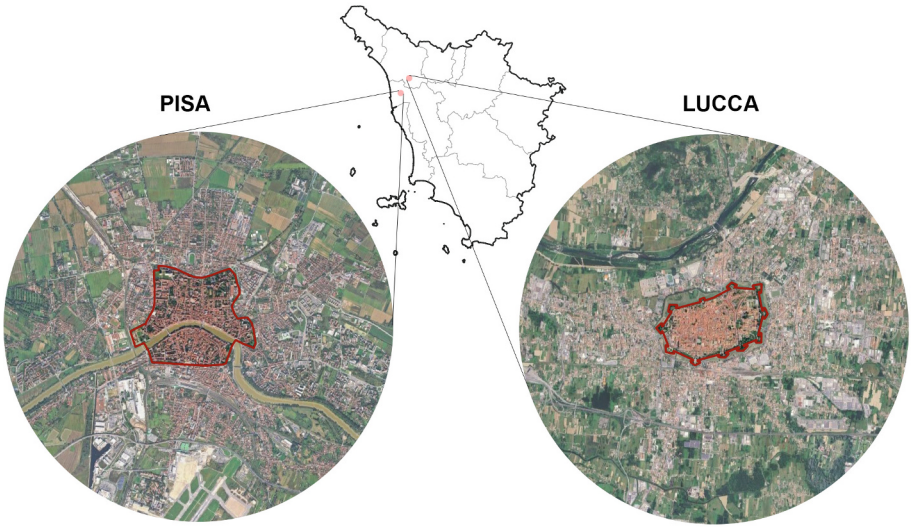


Fig. 1. Case studies

3 Methodological Framework

3.1 Data Acquisition: Ground-Truth vs. Digital Strata

The ground-truth database, which forms the basis of this study, was created specifically to overcome the limitations and privacy restrictions of official Italian business registers [11]. In line with the findings of Deri et al. [11] and Mara et al. [6], we conducted a systematic field mapping campaign focusing on commercial activities. Data collection was carried out independently by three surveyors on different weekdays during November 2025, strictly during normal working hours (approximately from 9:00 to 12:00 and from 16:00 to 19:00). The aim of this systematic approach was to obtain a representative snapshot of the active commercial landscape and to minimise discrepancies that might arise

from temporary shop closures. A comprehensive dataset was therefore created by integrating the collected records, which included the name, physical address, specific retail category and a unique alphanumeric identifier (ID) for each registered amenity. We then manually georeferenced the resulting dataset in QGIS, using Geoscopio orthophotos as the fundamental cartographic base.

In parallel with the field surveys, we extracted the OSM digital datasets by configuring geospatial queries using the QuickOSM plugin in QGIS to retrieve all point and multipolygon features tagged with 'amenity', 'shop', 'craft' and 'leisure'. The entire procedure was carried out in November 2025 to obtain data that were comparable with the surveys.

3.2 Spatial-Nominal Alignment

OSM entries do not contain complete address information; for this reason, aligning crowdsourced data with ground-truth records required a combined matching protocol based on geographical evidence and name information: an OSM entry was considered a 'match' if it fell within a 5-m spatial buffer of a real-world POI and shared a consistent business name. The choice of this radius was derived from two considerations: firstly, OSM data is generated by volunteers who mainly use smartphones [12] and the *Official US government information about the Global Positioning System (GPS) and related topics* states, based on research by Van Diggelen and Enge [13], that the GPS of smartphones is accurate on average within a 4.9 m radius under open sky. Furthermore, a test was previously conducted on the real database to evaluate the average distance between each business and its nearest neighbor. The Nearest Neighbour Index was found to be 0.075 which confirms the clustering of agglomerated points along shopping streets; the Observed Mean Distance presented a value of 10.31 m: the choice was therefore to fall within a 5-m buffer to prevent the search areas from overlapping with the adjacent activity.

Once matched, the OSM node inherited the corresponding real-world ID. If these criteria were not met, we assigned the unmatched OSM element a new, distinct unique identifier to formally record the absence of a corresponding real-world entity.

To analyse the nature of the data inconsistencies, we isolated and classified unmatched or problematic OSM entries into specific error types using a dedicated "mismatch" attribute:

- **Relocated/mislocated**: assigned when identical or equivalent business names appear in both datasets, but their spatial distance exceeds the 5-m threshold. In this case, the digital entry is linked to the real-world element via the same ID. This highlights the semantic presence on the one hand, whilst explicitly flagging the positional inaccuracy on the other, which likely stems from geocoding errors or the actual relocation of the business.
- **Replaced/inconsistent**: used when a digital POI shares the geographical location of a detected real entry but has a completely different name. This status reflects actual business turnover over time, as well as significant discrepancies in naming conventions.
- **Redundant**: when a single real-world activity is mapped multiple times in crowdsourced data (this can occur for both matching and non-matching records).

- **Excess**: reserved for 'ghost' web entries that have no corresponding real-world activity anywhere in the surveyed urban area.

3.3 The Meso-Category Taxonomy

The crux of this study lies in its approach to functional categorisation. Although useful for macro-scale models, the generic classifications often used in the literature obscure fundamental operational differences that are essential for micro-urban planning [1, 14]. To address this, Mara et al. [6] proposed an intermediate-level categorisation into 23 meso-categories to bridge the gap between overly generic classifications and the highly granular tags that typically describe geospatial data [15, 16]. By translating both the raw OSM data – via key-value tags – and the ground-truth observations into these 23 categories, it was possible to effectively demonstrate the platform's ability to accurately represent and maintain specific commercial sectors (Table 1). From a methodological point of view it is important to underline that, if an OSM node has been downloaded multiple times starting from different tags (e.g. shop = bakery and amenity = cafe which is a very common situation in the Italian context where hybrid premises with bar-pastry shops are frequent), each node was assigned to a single meso-category based on the specific tag with which it was downloaded. This procedure allowed us to avoid tag conflicts or multi-tag features. Next, during the matching phase and in particular during the assignment of the 'redundant' mismatch to one of the nodes, the meso-category was also considered to avoid distorting the Accuracy_cat results.

Table 1. Excerpt of the taxonomic alignment framework: correspondence between raw OSM tags (key = value), in-situ ground-truth labels, and the unified meso-categories used in this study

Meso-category	OpenStreetMap	Ground-truth retail category
Healthcare supplies and pharmacies	hearing_aids, medical_supply, pharmacy, herbalist, (…)	Farmacie (pharmacies), Parafarmacie (parapharmacies), (…)
Fuel	fuel	Carburanti (fuels)
Clothing and accessories	boutique, clothes, fashion, bag, fashion_accessories, (…)	Abbigliamento (clothing), Calzature (footwear), (…)
Private agencies	funeral_directors, travel_agency, storage_rental, money_lender, (…)	Agenzie immobiliari (real estate agencies), Agenzia viaggi (travel agency), (…)
Furniture and home furnishings	antiques, bathroom_furnishing, fireplace, bed, doors, (…)	Mobili (furniture), Cucine (kitchens), Letti (beds), (…)
Vehicles	atv, car, caravan, motorcycle, car_parts, fuel (shop), (…)	Autofficina (auto repair shop), Carrozzerie (car body shops), Autonoleggio (car rental), (…)

(*continued*)

Table 1. (*continued*)

Meso-category	OpenStreetMap	Ground-truth retail category
Workshops and crafts	agricultural_engines, atelier, bag_repair, bakery, (…)	Idraulici (plumbers), Fabbri (locksmiths), Falegnami (joiners), (…)
Home care	houseware, household_linen, dry_cleaning, (…)	Lavanderie (laundries), Casalinghi (housewares), (…)
Personal care	beauty, hairdresser, hairdresser_supply, massage, (…)	Parrucchieri (hairdressers), Estetista (beautician), (…)
Hardware and building supplies	hardware, key_cutter, insulation	Ferramenta (hardware store), Materiali edili (building materials), (…)
Electrical devices and sanitary equipments	electrical, security, energy (shop), hvac	Elettricisti (electricians), Materiale elettrico (electrical supplies), (…)
Bookshops, stationery and newsagents	books, paint, craft, stationery, copyshop, newsagent, kiosk	Librerie (bookstores), Cartolerie (stationery shops), (…)
Tobacco	lottery, tobacco	Tabaccherie (tobacco shops)
Photographic and optical equipments	photo_booth, camera, photo, photographer, photographic_laboratory, optician	Negozio di ottica (optical shop), Fotografia (photography), (…)
Miscellaneous items	cannabis, e-cigarette, gift, party, pyrotechnics, (…)	Articoli regalo (gift items), Bomboniere (party favours), (…)
Animals and plants	animal_boarding, animal_breeding, animal_training, (…)	Negozi di animali (pet shops), Fiorai (florists), Toelettatura animali (pet grooming), (…)
Technology and digital	electronics, radiotechnics, computer, printer_ink, (…)	Negozi di telefonia (mobile phone shops), Elettrodomestici (home appliances), (…)
Shopping malls	mall, general, wholesale, department_store	
Supermarkets	convenience, supermarket, variety_store, (…)	Supermercati (supermarkets), Minimarket (mini market / convenience store), (…)
Food and beverages	butcher, seafood, alcohol, beverages, water, wine, (…)	Macellerie (butcher shops), Pescherie (fish shops), Frutta e verdura (greengrocers), (…)

(*continued*)

Table 1. (*continued*)

Meso-category	OpenStreetMap	Ground-truth retail category
Restaurant services	food_court, restaurant	Ristoranti (restaurants), Pizzerie (pizzerias), Trattorie (traditional Italian restaurants), (…)
	fast_food, tortilla	Fast food (fast food restaurants), Kebab (kebab restaurants), (…)
Bars and clubs	bar, cafe, ice_cream, internet_cafe, pastry, (…)	Bar (bars), Caffetteria (cafés), Gelaterie (ice cream shops), (…)
Betting and gaming centres	casino, adult_gaming_centre, bookmaker, gambling, (…)	Agenzia scommesse (bookmaker), Sale gioco (arcades), (…)

3.4 Quantitative Evaluation

Using the unique ID as a join field, the two datasets were joined to generate a unified table comprising: (i) real-world activities successfully matched with OSM features; (ii) unmatched real-world activities; and (iii) all unmatched OSM features. This integrated table enabled a systematic comparison of the databases using the metrics defined in Table 2.

Following an overall assessment of OSM's reliability using the 'Amount' and 'Accuracy' metrics, we disaggregated the analysis by mesocategory to investigate two specific dimensions. Firstly, we examined how accurately real-world categories are reflected within the OSM database, with the aim of identifying potential trends relating to the specific type of goods sold or services provided. Subsequently, we assessed the accuracy of the categorisation of web entries to determine whether the degree of representativeness observed in the first phase translated into correct intrinsic categorisation within the OSM platform.

Table 2. Metrics analysed for assessing OSM reliability

Measure	Explanation	Equation
Amount	Assessment of the apparent data over- or under-estimation within the web platform	$A = \frac{N_{OSM}}{N_{real}}$
Accuracy	Measure of the actual digital platform capability to map existing ground-truth features…	$AC = \frac{M_{OSM}}{N_{real}}$
Accuracy_1	…including the relocated/mislocated entries	$AC_1 = \frac{M_{OSM} + M_{OSM_rel}}{N_{real}}$
Accuracy_cat	The proportion of successfully matched web entries that exhibit the correct functional categorisation	$AC_C = \frac{M_{OSM_cat}}{N_{real}}$

4 Results and Analytical Discussion

4.1 Overall Assessment

The initial quantitative assessment focused on the overall volume of data and its spatial validity. The on-site mapping campaign verified the existence of 1,236 active commercial entities within the historic walls of Pisa and 792 in Lucca. By contrast, the extraction from OSM yielded a significantly lower volume, with 837 nodes for Pisa and 407 for Lucca. An assessment of the Amount metric (OSM entries extracted relative to the total number of real-world entities) reveals a significant shortfall in data completeness: in Pisa, the dataset captures only 68% (A = 0.68) of the actual commercial fabric, whilst in Lucca coverage stands at 51% (A = 0.51).

Aware that volume alone is a misleading metric, we analysed the Accuracy – defined as the proportion of OSM entries that correctly corresponded to a physical commercial business (see Table 2) – which revealed another vulnerability: in both historic centres, Accuracy dropped significantly, reaching just 50% in Pisa (AC = 0.50) and 30% in Lucca (AC = 0.30). Even when increasing the matching tolerance to include incorrectly placed entries ('relocated/mislocated') as positive matches, the overall Accuracy scores show no substantial improvement (AC_1 = 0.51 for Pisa and AC_1 = 0.32 for Lucca).

The reasons behind this sharp decline from the initial Amount to the actual Accuracy become clear when examining the types of mismatches. Both case studies reveal an almost total absence of duplicate elements, with only 1% of OSM features extracted in Pisa and 3% in Lucca classified as 'redundant'. Similarly, no substantial differences are observed regarding non-existent digital records: 14% of OSM entries in Pisa and 11% in Lucca are labelled 'excess'. Conversely, records with completely divergent names show a different trend, with 'replaced/inconsistent' entries accounting for 9% in Pisa and 22% in Lucca.

Thus, the OSM dataset proves to be positionally robust, free from gross spatial errors and essentially devoid of extraneous elements. The 'noisy' portion of the data can in fact be attributed either to semantic errors during data entry into the platform or, more seriously, to a systemic temporal vulnerability inherent in participatory mapping when applied to these complex urban environments. This latter hypothesis is supported by the morphological and functional characteristics shared by the two case studies (Sect. 2): within the dense alleyways of these medieval cities, volunteers' mapping efforts structurally struggle to keep pace with the rapid turnover of small shops [17].

Ultimately, the main problem is a significant time lag rather than a mere spatial inaccuracy, which results in a data landscape cluttered with the 'digital ghosts' of businesses that have since closed down or become obsolete.

4.2 The Meso-Category Assessment

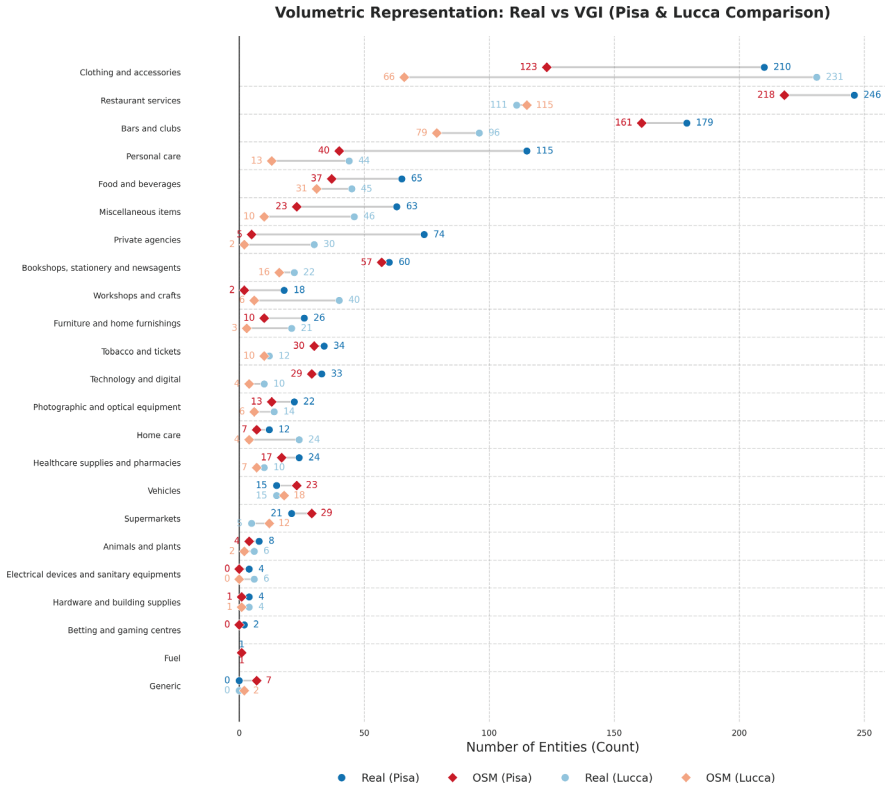


Fig. 2. Dumbbell plot illustrating the volumetric difference between OSM entries and ground-truth activities within the historical centres of Pisa and Lucca. The length of the grey line represents the magnitude of the gap between the two datasets: specifically, a circle positioned to the left of the diamond indicates an overrepresentation of entries in the OSM database, whereas a circle to the right denotes an underrepresentation

A detailed analysis at the meso-category level reveals a thematic imbalance in the mapping patterns of Volunteered Geographic Information (VGI), which is immediately evident when examining the Amount metric. In both Lucca and Pisa, the dataset shows widespread under-representation regardless of the type of shop. As illustrated in Fig. 2, the web data markers (diamonds) are consistently lower than the ground-truth counts (circles); overestimates occur only in a few isolated cases, such as the 'Supermarkets' and 'Vehicles' categories, which display Amount values greater than 1.0 in both contexts. The over-representation of these specific businesses within the analysed sample likely reflects a time lag in digital curation, linked in particular to the delayed updating or failure to remove obsolete data. This evidence suggests that the rapid physical turnover of these commercial businesses is not accompanied by an equally rapid maintenance of the web database, leading to an accumulation of obsolete records.

The categories with the highest coverage (A > 0.80) are 'Tobacco and tickets', 'Restaurant services' and 'Bars and clubs' – that is, highly frequented establishments that are vital both for the local social fabric and for the tourism economy. These patterns are strongly consistent with the academic literature on Points of Interest (POI) bias and thematic heterogeneity [18, 19], according to which VGI contributors tend to prioritise highly visible, heavily trafficked or socially significant urban amenities.

As already suggested in Sect. 4.1, relying solely on coverage indicators gives a false impression of completeness, as their interpretation must be supplemented by spatial reliability metrics. Accuracy values exceeding 0.50 confirm the results mentioned above: 'Restaurant services', 'Supermarkets', 'Bars and clubs', and 'Healthcare supplies' consistently emerge as the categories represented with the greatest precision in both case studies. Conversely, a stark contrast between high volumetric representation and poor spatial-semantic correspondence is exemplified by the categories 'Tobacco and tickets' and 'Vehicles' in Lucca: despite having significant volumes (A = 0.83 and A = 1.20, respectively), their accuracy drops to a mere 0.25 and 0.33 (see Table 3). This highlights that, in addition to a surplus of records, even the data points that would numerically overlap with real-world entities exhibit a serious qualitative deficit.

More broadly, particularly low Accuracy scores emerge across a clearly defined group of categories involving *comparison goods retail* [20]: sectors such as 'Clothing and accessories', 'Electrical devices and sanitary equipment', 'Hardware and building supplies', and 'Furniture and home furnishings' consistently show poor reliability in both case studies. This trend suggests that whilst VGI contributors readily map and update social or tourist-oriented services, traditional retail shops attract significantly less digital engagement.

The fact that Pisa outperforms Lucca in terms of overall data quality across most categories suggests a potentially more active local mapping community. Despite this, in both urban centres, expanding the spatial matching tolerance to include 'relocated/mislocated' entries (Accuracy_1) yields negligible improvements, confirming that the main issue within these historic districts is not simple GPS inaccuracy, but rather a complex combination of a complete lack of data and significant temporal lag (Table 3).

Mismatch Typologies. A detailed analysis of the types of mismatches provides key insights into the mechanisms underlying the deterioration of VGI data in the case studies examined, confirming that geometric imprecision is largely secondary to poor temporal and lifecycle management (Fig. 3).

Firstly, in both Lucca and Pisa, the vast majority of mapping anomalies fall into the 'excess' and 'replaced/inconsistent' categories. On closer inspection, however, it emerges that 'excess' records (surplus or ghost nodes) predominate in Pisa, whilst phenomena linked to actual business turnover or serious discrepancies in nomenclature ('replaced/inconsistent') are more widespread in Lucca. This suggests that in Pisa there is a prevailing failure to remove obsolete records, highlighting the locally additive nature of crowdsourced mapping: volunteer contributors demonstrate a strong proactive commitment to creating new spatial entities, but systematically neglect the subtractive maintenance required to remove closed or obsolete businesses. The city's strong focus on tourism and academia likely exacerbates this trend, particularly within highly dynamic

categories such as ‘Bars and clubs’, ‘Bookshops, stationery and newsagents’, ‘Clothing and accessories’, ‘Food and beverages’, and ‘Vehicles’.

On the other hand, Lucca appears to be affected by a more general lack of data updates, with a pool of local contributors that is less active and shows little inclination to add or update entries. Here too, the city’s economy, which is heavily tourism-oriented, may help to explain this phenomenon, which particularly affects categories such as ‘Bars and clubs’, ‘Clothing and accessories’, and ‘Restaurant services’.

Ultimately, in both urban contexts, the dataset functions more as a cumulative historical record than as a real-time synchronised digital twin.

Furthermore, the distribution of ‘redundant’ records is concentrated almost exclusively in the macro-domain of food service and in the ‘Vehicles’ category. Rather than mere random duplication, this likely reflects the functional complexity and hypervisibility of these venues. Indeed, very popular locations are often mapped by multiple independent users who may represent the same physical entity using different tags that nevertheless refer to the same functional sphere.

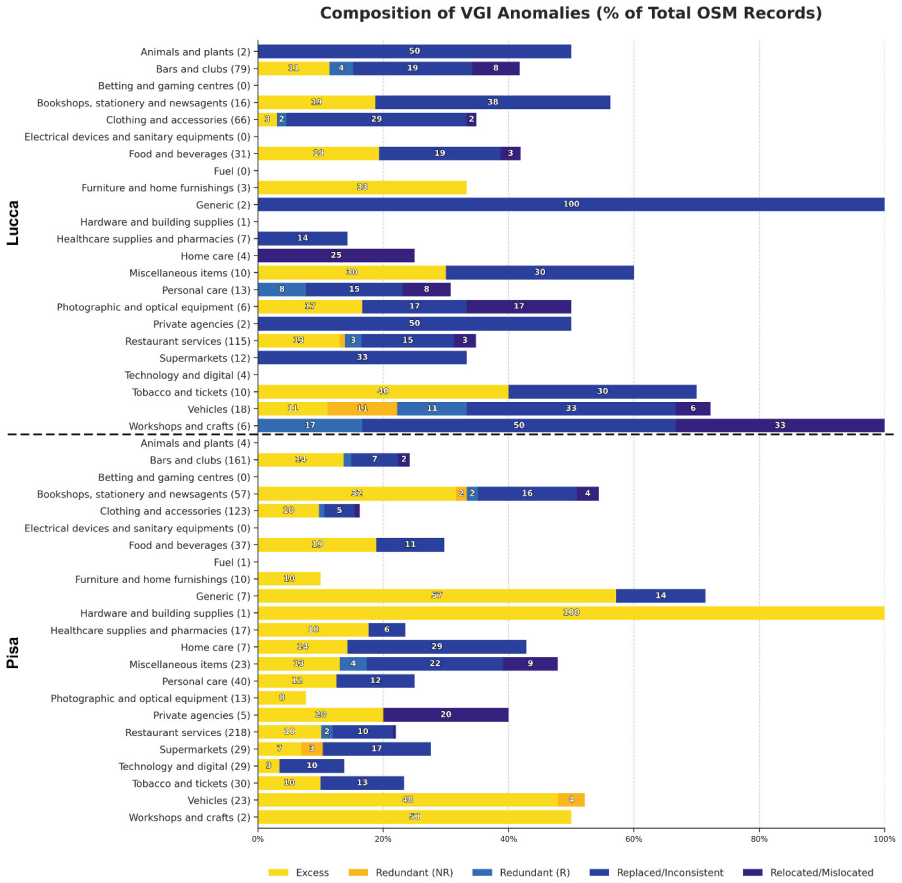


Fig. 3. Bar chart showing mismatch typologies by meso-categories. Data are normalized to the OSM values (shown in parentheses).

Table 3. Reliability assessment metrics for OSM dataset

Area	Meso-category	N_{real}	N_{OSM}	A	M_{OSM}	AC	M_1*	AC_1
Lucca	Animals and plants	6	2	0.33	1	0.17	1	0.17
	Bars and clubs	96	79	0.82	52	0.54	58	0.60
	Bookshops, stationery and newsagents	22	16	0.73	5	0.23	5	0.23
	Clothing and accessories	231	66	0.29	38	0.17	39	0.17
	Electrical devices and sanitary equipments	6	0	0.00	1	0.17	1	0.17
	Food and beverages	45	31	0.69	18	0.40	19	0.42

(*continued*)

Table 3. (*continued*)

Area	Meso-category	N_{real}	N_{OSM}	A	M_{OSM}	AC	M_1*	AC_1
	Furniture and home furnishings	21	3	0.14	2	0.10	2	0.10
	Generic	0	2		0		0	
	Hardware and building supplies	4	1	0.25	1	0.25	1	0.25
	Healthcare supplies and pharmacies	10	7	0.70	6	0.60	6	0.60
	Home care	24	4	0.17	5	0.21	6	0.25
	Miscellaneous items	46	10	0.22	8	0.17	8	0.17
	Personal care	44	13	0.30	9	0.21	10	0.23
	Photographic and optical equipment	14	6	0.43	4	0.29	5	0.36
	Private agencies	30	2	0.07	1	0.03	1	0.03
	Restaurant services	111	115	1.04	71	0.64	75	0.68
	Supermarkets	5	12	2.40	5	1.00	5	1.00
	Technology and digital	10	4	0.40	2	0.20	2	0.20
	Tobacco and tickets	12	10	0.83	3	0.25	3	0.25
	Vehicles	15	18	1.20	5	0.33	6	0.40
	Workshops and crafts	40	6	0.15	2	0.05	4	0.10
Pisa	Animals and plants	8	4	0.50	4	0.50	4	0.50
	Bars and clubs	179	161	0.90	135	0.75	138	0.77
	Betting and gaming centres	2	0	0.00	0	0.00	0	0.00
	Bookshops, stationery and newsagents	60	57	0.95	24	0.40	26	0.43
	Clothing and accessories	210	123	0.59	103	0.49	104	0.50
	Electrical devices and sanitary equipments	4	0	0.00	1	0.25	1	0.25
	Food and beverages	65	37	0.57	25	0.39	25	0.39
	Fuel	1	1	1.00	1	1.00	1	1.00
	Furniture and home furnishings	26	10	0.39	6	0.23	6	0.23
	Generic	0	7		0		0	
	Hardware and building supplies	4	1	0.25	0	0.00	0	0.00
	Healthcare supplies and pharmacies	24	17	0.71	13	0.54	13	0.54

(*continued*)

Table 3. (*continued*)

Area	Meso-category	N_{real}	N_{OSM}	A	M_{OSM}	AC	M_1*	AC_1
	Home care	12	7	0.58	5	0.42	5	0.42
	Miscellaneous items	63	23	0.37	24	0.38	26	0.41
	Personal care	115	40	0.35	31	0.27	31	0.27
	Photographic and optical equipment	22	13	0.59	11	0.50	11	0.50
	Private agencies	74	5	0.07	3	0.04	4	0.05
	Restaurant services	246	218	0.89	163	0.66	164	0.67
	Supermarkets	21	29	1.38	13	0.62	13	0.62
	Technology and digital	33	29	0.88	21	0.64	21	0.64
	Tobacco and tickets	34	30	0.88	21	0.62	21	0.62
	Vehicles	15	23	1.53	10	0.67	10	0.67
	Workshops and crafts	18	2	0.11	4	0.22	4	0.22

*M_1 denotes the sum $M_{OSM} + M_{OSM\ _\ rel}$

Semantic Accuracy. Finally, assessing the correctness of the classification using the Accuracy_cat metric confirms that, despite the aforementioned volumetric inflation and spatial gaps, when a physical entity is successfully matched with an OSM node, the accuracy of its categorisation is remarkably high (Fig. 4). In both case studies, Accuracy_cat often records scores between 0.80 and 1.0 (14/23 categories in Lucca and 15/23 in Pisa), indicating a dichotomy in the quality of VGI: whilst volunteer mappers struggle to capture the full extent and update the spatial layout of the ever-evolving urban commercial fabric, their semantic classification of the nodes they actually choose to map appears highly reliable.

It is important to emphasise that this high level of semantic agreement is structurally facilitated by the meso-level categorisation framework adopted. It is, in fact, a reference that is neither overly granular nor too generic, and which effectively intercepts gross conceptual errors. It also allows for the harmonisation of internal tagging inconsistencies inherent to VGI – which frequently stem from the heterogeneous linguistic, cultural and socio-spatial contexts that guide contributors in the recognition and classification of local amenities [6]. Within this framework, it emerges that categories characterised by universally recognised and unambiguous physical profiles – such as 'Supermarkets', 'Healthcare supplies and pharmacies', 'Clothing and accessories', 'Vehicles', and 'Animals and plants' – consistently achieve near-perfect categorisation scores in both Lucca and Pisa; for these services, the semantic translation from a physical shopfront to a digital tag is likely straightforward for a volunteer mapper.

Conversely, categorisation accuracy decreases in two specific scenarios. Firstly, within the highly mapped macro-domain of food service (e.g. 'Bars and clubs', 'Food and beverages'), Accuracy_cat shows a slight variation, generally ranging between 0.60 and 0.95. This reflects the actual physical and functional hybridity of these venues, where volunteers correctly identify the broader function related to food but may diverge

conceptually from the ground-truth classification. Secondly, the lowest semantic accuracies – often ranging between 0.25 and 0.60 – are systematically recorded in conceptually broad or highly specialised sectors, such as 'Miscellaneous items' and 'Workshops and crafts'. In these cases, the lack of standardised visual cues and the ambiguous nature of the services make it challenging for the crowdsourcing community to classify the physical entity consistently. Ultimately, whilst extreme failures (such as 0% in 'Electrical devices') are simply statistical artefacts resulting from very small spatial matching sets, the broader trend confirms that the semantic reliability of VGI is directly proportional to the ontological clarity of the mapped object.

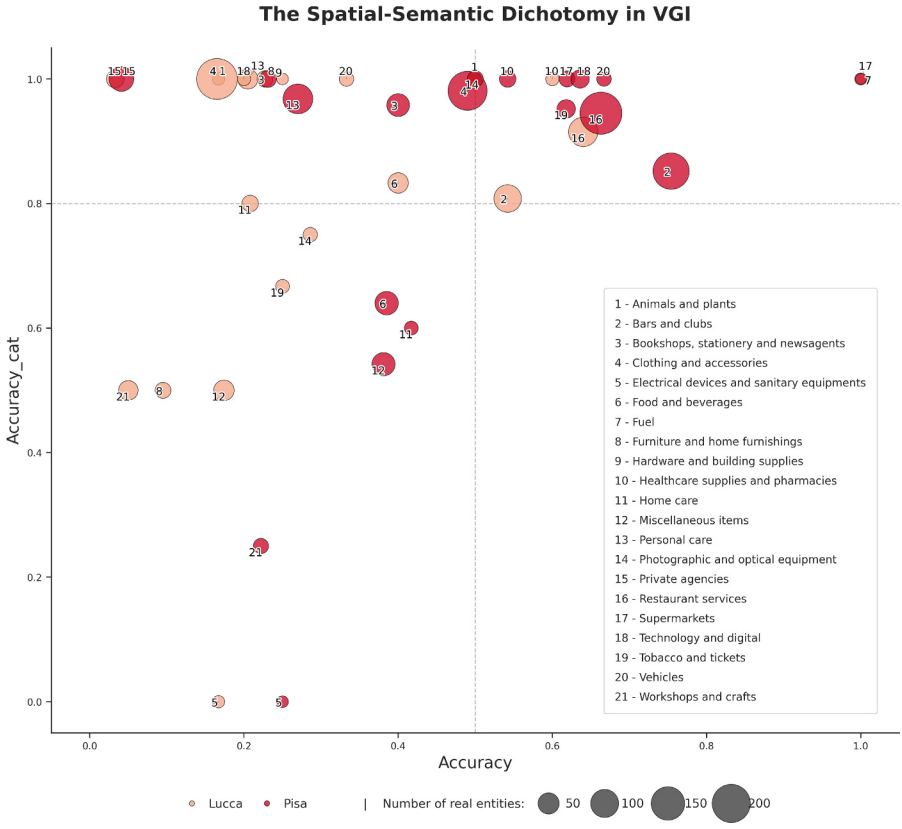


Fig. 4. Bubble scatter plot illustrating spatial accuracy against categorization reliability. The bubble size represents the absolute count of real-world activities. The predominant clustering within the upper section of the plot indicates that OSM entries are generally correctly categorized, regardless of their varying spatial accuracy values

5 Conclusions

As urban planning paradigms increasingly rely on granular spatial analyses to model proximity-based environments, the uncritical use of crowdsourced datasets entails significant methodological risks. This study, through a ground-truth validation in the historic centres of Pisa and Lucca, highlights the structural and behavioural limitations of OpenStreetMap in capturing the highly changeable commercial fabric of these complex urban centres. In particular, three key findings emerge from this research.

Firstly, VGI is intrinsically shaped by the thematic heterogeneity and bias of Points of Interest (POI). In both contexts, the mapping community systematically prioritises highly visible, socially relevant or tourism-oriented services (e.g., the food-related and mobility sectors), whilst neglecting ordinary, localised retail trade.

Secondly, the data reveal a stark contrast between spatial completeness and semantic reliability. The study shows that traditional commerce suffers from severe under-mapping attributable to a genuine lack of data rather than geometric imprecision (even relaxing spatial tolerances fails to bridge the gap), but it is also evident that successfully mapped nodes are classified with considerable accuracy.

Thirdly, and this is the most critical point, the analysis of the types of mismatches allows to hypothesize that OSM suffers from a serious lack of subtraction-based maintenance in both case studies. The prevalence of 'excess' and 'replaced/inconsistent' mapping anomalies confirms that crowdsourced platforms struggle to keep pace with the rapid turnover in the retail sector: volunteer contributors actively map new establishments, but fail to systematically record urban commercial churn. Consequently, the platform seems to function as an additive and cumulative historical record rather than as a real-time synchronised digital twin, leaving the digital landscape littered with 'digital ghosts' of relocated or closed businesses. To confirm this hypothesis, it is possible to conduct a subsequent in-depth analysis by verifying the timestamps of individual OSM records. At the same time, separating the contributions of never-mapped entities from stale-but-once-correct entities – which in this study were not isolated because they were marginal to the specific research question – would allow to outline different remediation strategies for web platform managers.

While the empirical analysis of this study is grounded in the cities of Pisa and Lucca, the general methodology extends beyond these specific geographical boundaries. The manual mapping and matching process means that the protocol in its current form is difficult to reproduce over large contexts, but future developments could automate the process to make it replicable at scale. In this way it would be possible to broaden the scope of investigation to include even more morphologically and functionally complex case studies, also in international contexts, to ascertain whether the trends and biases observed in Pisa and Lucca are intrinsic to the VGI platform itself rather than being purely place-based. Furthermore, the implementation of Artificial Intelligence-based tools will be crucial to systematically eliminate obsolete nodes and align digital layers with the relentless pace of urban evolution, ultimately bridging the gap between historical crowdsourced records and real-time synchronised digital twins.

In conclusion, although OSM remains an invaluable resource for mapping static infrastructure networks, its use in the dynamic analysis of urban retail must be approached with extreme caution [7, 21]. If unverified crowdsourced data is fed, for

example, into '15-min city' algorithms, the resulting models could overestimate the availability of hospitality services and drastically underestimate the retail services essential for daily life, leading to distorted accessibility metrics.

Acknowledgments. This work was part of the emc2 project, funded by ANR (France), FFG (Austria), MUT (Italy) and Vinnova (Sweden) under the Driving Urban Transition Partnership, which has been funded by the European Commission.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Moreno, C., Allam, Z., Chabaud, D., Gall, C., Pratlong, F.: Introducing the "15-Minute City": sustainability, resilience and place identity in future post-pandemic cities. Smart cities. **4**(1), 93–111 (2021)
2. Allam, Z., Bibri, S.E., Chabaud, D., Moreno, C.: The theoretical, practical, and technological foundations of the 15-minute city model: proximity and its environmental, social and economic benefits for sustainability. Energies. **15**(16), 6042 (2022)
3. Tahmasbi, B., Mansourianfar, M.H., Haghshenas, H., Kim, I.: Multimodal accessibility-based equity assessment of urban public facilities distribution. Sustain. Cities Soc. **49**, 101633 (2019)
4. Anselmi, C., Mara, F., Deri, F.: Big-data reliability in informing the 15-Min City modelling: a comparison between OpenStreetMap and Google maps. In: International Conference on Computational Science and its Applications, pp. 325–341. Springer Nature Switzerland, Cham (2025)
5. Herfort, B., Lautenbach, S., Porto de Albuquerque, J., Anderson, J., Zipf, A.: A spatio-temporal analysis investigating completeness and inequalities of global urban building data in OpenStreetMap. Nature. Communications. **14**(1), 3985 (2023)
6. Mara, F., Deri, F., Anselmi, C.: Mismapped cities: the risks of using big-data in urban planning. In: 13th International Conference on Innovation in Urban and Regional Planning, Cham, Springer Nature, Switzerland (forthcoming)
7. Brovelli, M.A., Minghini, M., Molinari, M., Mooney, P.: Towards an automated comparison of OpenStreetMap with authoritative road datasets. Trans. GIS. **21**(2), 191–206 (2017)
8. Han, C., Lu, B., Zheng, J., Yu, D., Zheng, S.: Research on multiscale OpenStreetMap in China: data quality assessment with EWM-TOPSIS and GDP modeling. Geo-spat. Inf. Sci. **28**(3), 1316–1340 (2025)
9. Zhang, J., Yu, Z., Li, Y., Wang, X.: Uncovering bias in objective mapping and subjective perception of urban building functionality: a machine learning approach to urban spatial perception. Land. **12**(7), 1322 (2023)
10. Gehl, J.: Cities for People. Island press, Washington (2010)
11. Deri, F., Mara, F., Anselmi, C.: Crowdsourced data for urban planning: a critical evaluation of OpenStreetMap accuracy and completeness. In: 25th International Conference on Computational Science and its Applications. Cham, Springer Nature, Switzerland (2025)
12. Fonte, C.C., et al.: Assessing VGI data quality. In: Mapping and the Citizen Sensor. Ubiquity Press (2017)
13. Van Diggelen, F., Enge, P.: The world's first GPS MOOC and worldwide laboratory using smartphones. In: Proceedings of the 28th International Technical Meeting of the Satellite Division of the Institute of Navigation, pp. 361–369 (2015)

14. Papadopoulos, E., Sdoukopoulos, A., Politis, I.: Measuring compliance with the 15-minute city concept: state-of-the-art, major components and further requirements. Sustain. Cities Soc. 104875 (2023)
15. Place Types: Google Maps Platform Documentation. Available online: https://developers.google.com/maps/documentation/places/web-service/place-types#table-a. Accessed Mar 2026
16. OpenStreetMap-Map features. https://wiki.openstreetmap.org/wiki/Map_features. Accessed Mar 2026
17. Meltzer, R., Capperis, S.: Neighbourhood differences in retail turnover: Evidence from new York City. Urban Stud. **54**(13), 3022–3057 (2017)
18. Haklay, M.: How good is volunteered geographical information? A comparative study of OpenStreetMap and ordnance survey datasets. Environ. Plann. B. Plann. Des. **37**(4), 682–703 (2010)
19. Psyllidis, A., et al.: Points of interest (POI): a commentary on the state of the art, challenges, and prospects for the future. Comput. Urban Sci. **2**(1), 20 (2022)
20. Dolega, L., Lord, A.: Exploring the geography of retail success and decline: a case study of the Liverpool City region. Cities. **96**, 102456 (2020)
21. Ferster, C., Fischer, J., Manaugh, K., Nelson, T., Winters, M.: Using OpenStreetMap to inventory bicycle infrastructure: a comparison with open data from cities. Int. J. Sustain. Transp. **14**(1), 64–73 (2020)