Overture POI data for the United Kingdom: a comprehensive, queryable open data product

Patrick Ballantyne ⬤

University of Liverpool

p.ballantyne@liverpool.ac.uk

Cillian Berragan ⬤

University of Liverpool

c.berragan@liverpool.ac.uk

2023-10-27

Abstract

Point of Interest data that is comprehensive, globally-available and open-access, is sparse, despite being important inputs for research in a number of application areas. New data from the Overture Maps Foundation offers significant potential in this arena, but accessing the data relies on computational resources beyond the skillset and capacity of the average researcher. In this article, we provide a processed version of the Overture places (POI) dataset for the UK, in a fully-queryable format, and provide accompanying code through which to explore the data, and generate other national subsets. In the article, we describe the construction and characteristics of the dataset, before considering how reliable it is (locational accuracy, attribute comprehensiveness), through direct comparison with Geolytix supermarket data. This dataset can support new and important research projects in a variety of different thematic areas, and foster a network of researchers to further evaluate its advantages and limitations.

Keywords points of interest • overture • amazon web services

1 Introduction

Point of Interest (POI) data is an invaluable source of information, acting as a key input to much of the research that has, and continues to be generated in urban analytics and city science. These data provide key locational attributes about a broad variety of social, environmental and economic phenomena, including historical landmarks, parks, hospitals and retailers, and have been vital sources of data for different applications, including health (Green et al. 2018; Hobbs et al. 2019), urban mobility (Graells-Garrido et al. 2021; Jay et al. 2022), retail and location analysis (Ballantyne et al. 2022), transportation (Owen, Arribas-Bel, and Rowe 2023; Credit 2018), and many others. However, a major challenge when working with POI data relates to the coverage and comprehensiveness of these datasets (Ballantyne et al. 2022; Zhang and Pfoser 2019). By this we mean how much the chosen source(s) of POI data restricts the analyses to specific cities or regions (i.e., coverage), and the attributes and characteristics that are provided for each POI (i.e., comprehensiveness).

Many POI datasets offer a high level of global coverage and availability, such as OpenStreetMap. However there are problems when considering the coverage and comprehensiveness of OpenStreetMap data at finer spatial resolutions and in areas with less contributors (Haklay 2010), as well as in less developed countries (Mahabir et al. 2017). Similarly, datasets like OpenStreetMap often contain inconsistent attributes for economic activities like retail stores and leisure (Zhang and Pfoser 2019; Ballantyne et al. 2022). Some POI datasets exist which fill this gap, such as the Ordnance Survey 'Points of Interest' data product, which provides a more comprehensive database of economic activities (Haklay 2010), but is not openly-available. Other data providers have democratised access to comprehensive POI datasets such as SafeGraph and the Local Data Company, however these datasets exhibit poor global coverage of non-branded POIs (SafeGraph), and a lack of comprehensive coverage in the UK (Dolega et al. 2021). As a result, there is a clear gap for data that can address some of these limitations, by providing an openly-available, comprehensive

and accurate source of POIs for the UK. In this article, we introduce readers to a processed version of the Overture Maps places (POI) dataset (Overture Maps Foundation 2023), which arguably provides a strong solution to many of these problems, and can facilitate groundbreaking urban analytics research in a number of different application areas.

2 Data

The data was accessed through the Overture Maps Foundation, which was set up as a collaborative venture to develop reliable, easy-to-use, and interoperable open map data (Overture Maps Foundation 2023). The foundation, which is steered by Amazon, Microsoft, Meta and TomTom, has developed a number of open data products including Buildings, Places, Transportation and Administrative Geographies, all of which are available at global scales and contain a detailed number of attributes (Overture Maps Foundation 2023). Users can access the data parquet files from the cloud using Amazon Athena, Microsoft Synapse or DuckDB, or download them locally. However, a specific challenge for urban analytics researchers and city scientists is that the majority will not have the data engineering skills to query these datasets from the cloud, and process the attributes in their nested JSON format. Furthermore, for those who want to download the files locally, they can be difficult to work with, as the full global places file is over 200 GB. Therefore, our aim is to provide a processed subset of the Overture places dataset for the UK, which bypasses these issues, and creates an open data product for use in research.

Overture hosts all data through Amazon Web Services (AWS), which enables a number of query end points to be used to download data subsets. The Overture data schema includes a bounding box structure column to enable efficient spatial SQL queries. To query POI data for the UK, a spatial SQL query was constructed using the DuckDB SQL engine and the UK bounding box, based on EPSG:27700. This query downloaded a GeoPackage file containing all POIs within the UK bounding box, totalling 1.34 GB. This file was then clipped to the administrative boundaries of the United Kingdom, to exclude non-UK places that appeared within the bounding box query.

As noted, many of the columns that provide metadata relating to POIs are represented in a nested JSON format (columns containing lists of lists), which are difficult to efficiently parse with traditional tabular data frame libraries. We therefore processed the following columns to ensure the data frame remained two-dimensional: Names, Category, Address and Brand. Following this processing, we spatially joined the 2021 census area geographies for England including Output Areas (OA), Lower layer Super Output Areas (LSOA), Middle layer Super Output Areas (MSOA), and the 2022 Local Authority Districts (LAD). For both Scotland and Northern Ireland, we spatially joined the 2011 Data Zone geographies. We also include the H3 (hexagons) addresses associated with each point for all resolutions between 1 and 9. The resulting dataset is a 358 MB GeoParquet file, hosted as part of a DagsHub data repository, and the final processed data file, comprising the Overture POI subset for the UK can be easily downloaded[1]. A list of attributes for the data product can be found in Table i (supplementary materials), and as a secondary output of this paper, an example workflow for how to extract Overture places for other study areas has also been produced[2].

3 Reliability Analysis - Retail Brands

To assess the reliability of Overture places, we compared them with the Geolytix Supermarket Retail Points dataset (Geolytix 2023), which is known to provide reliable information about supermarkets in the UK, and provides a useful 'ground-truth' dataset to test how well Overture represents economic activities. In particular, we examined how many of the Geolytix supermarkets are captured in Overture, the accuracy of the POI coordinates, and how complete the category/brand information is. Table 1 shows that the Overture data aligns well with the Geolytix data, with small differences across the three retailers (< 5%). Table 1 shows that there was a relatively low median distance (metres) between Overture points and their closest Geolytix point, evidencing a relatively high level of accuracy in terms of geographical positioning. This is an important attribute, as incorrect positioning of POI data can have dramatic implications for accessibility measurement (Green et al. 2018; Graells-Garrido et al. 2021), and urban boundary delineation (Ballantyne et al. 2022).

In terms of the comprehensiveness of the category and brand information, a large number of the Overture POIs contained missing values for categories or brands (Table 2), making filtering of the dataset to a specific retailer (e.g., Waitrose), slightly less simple. Table 2 displays the complexity of these issues, where different degrees of completeness are apparent when considering the source of the POI (Meta or Microsoft). This has strong implications for how Overture data can and should be used, especially for applications involving specific POI categories or brands. Whilst it is not impossible to extract a complete list of POIs for a retailer, through collective filtering of POI name, brand and categories to collect these features (see supplementary materials), users should be aware of the high level of attribute incompleteness for POIs extracted from Microsoft. Further reliability analysis is beyond the scope of this paper, but

---

[1]https://figshare.com/s/144265a705159c03c08f?file=42761512
[2]https://figshare.com/s/144265a705159c03c08f?file=42809656

Table 1: Reliability analysis of Overture compared with Geolytix retail points dataset.

Retailer	Geolytix count	Overture count	Average distance between points (m)
Waitrose	422	420	8.3
Spar	2,339	2,308	6.5
Tesco	2,840	2,753	6.2

Table 2: Overture attributes compared with Geolytix retail points dataset.

	Attribute incompleteness (%)			
	Category information		Brand information	
Retailer	Meta	Microsoft	Meta	Microsoft
Waitrose	100	N/A	23.33	N/A
Spar	0.18	100.00	11.63	100.00
Tesco	0.00	N/A	2.29	N/A

there is a clear need for further investigation into how well Overture places captures category and brand information for other non-retail POIs (e.g., GP practices, post offices).

4 Application - Mapping supermarkets in the UK

To demonstrate how this dataset can be used, an example workflow has been presented which reads in the UK processed version of Overture places, filters to a specific brand of supermarket, and then maps the distribution of these nationally (Figure 1). The purpose of these workflows is to illustrate how easy it is to work with this dataset, and the variety of different POI attributes that are stored within the dataset. Example workflows have been presented for both the Python[3] and R[4] programming languages, and utilise preferred packages for data manipulation and mapping (e.g., arrow, geopandas)

5 Conclusion

This paper presents a comprehensive, queryable open data product, which represents a processed UK national subset of the Overture places database. This new open data product makes Overture data more accessible for researchers, bypassing the need for advanced data engineering skills and large amounts of memory on which to store the complete database. The potential applications of this data product in a variety of different fields is highly significant (e.g., urban accessibility), given the evidence presented about the coverage, comprehensiveness and locational accuracy of this dataset. At a time where the retail sector is undergoing significant transformations in response to the cost-of-living crisis, such data can provide invaluable insights about the characteristics and performance of the sector (Ballantyne et al. 2022; Dolega et al. 2021), which has historically been a challenge due to the availability of suitable retailer data. However, there are inherent limitations to this dataset, which have been illustrated through direct comparison with Geolytix data. Users need to be cautious about how they are using this data, especially when the POIs they are using are largely sourced from Microsoft. However, it is our hope that by releasing this data into the open domain, a network of researchers will be fostered who can utilise this data for their own research questions, and critically evaluate how the Overture places database represents a variety of different social, economic and environmental activities.

6 Data Availability Statement

The UK Overture data product (anonymised for peer review) can be downloaded directly from Figshare: https://figshare.com/s/144265a705159c03c08f?file=42761512. The data product can be directly queried from the DagsHub repository, but for the purposes of anonymous peer review, this has not been included in the paper.

References

Ballantyne, Patrick, Alex Singleton, Les Dolega, and Kevin Credit. 2022. "A Framework for Delineating the Scale, Extent and Characteristics of American Retail Centre Agglomerations." *Environment and Planning B: Urban Analytics and City Science* 49 (3): 1112–28.

---

[3]https://figshare.com/s/144265a705159c03c08f?file=42809500
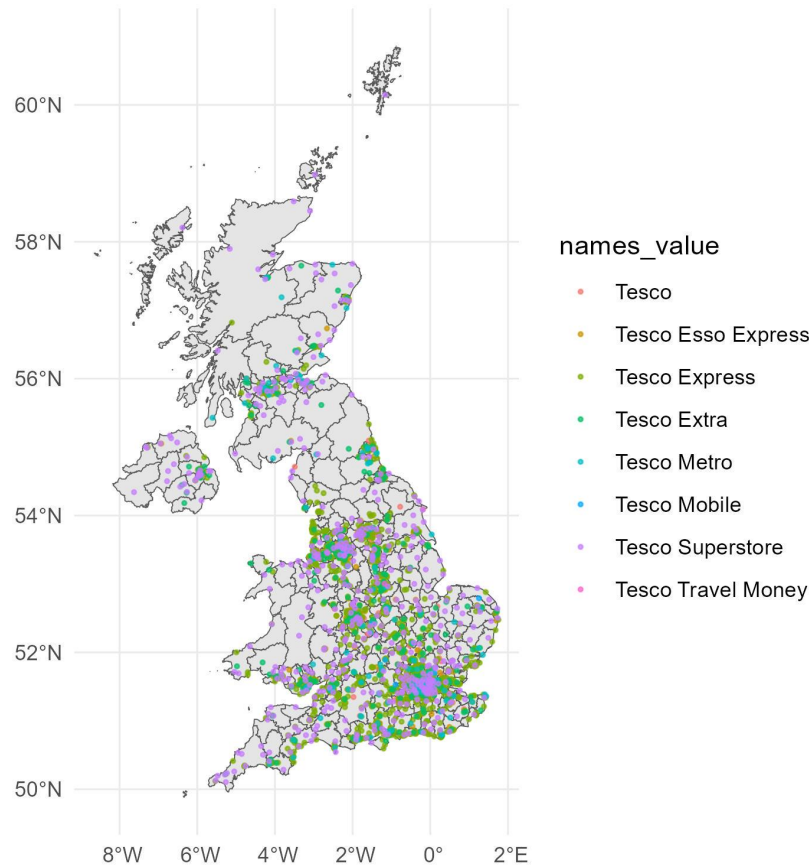[4]https://figshare.com/s/144265a705159c03c08f?file=42809452

Figure 1

Credit, Kevin. 2018. "Transit-Oriented Economic Development: The Impact of Light Rail on New Business Starts in the Phoenix, AZ Region, USA." *Urban Studies* 55 (13): 2838–62.

Dolega, Les, Jonathan Reynolds, Alex Singleton, and Michalis Pavlis. 2021. "Beyond Retail: New Ways of Classifying UK Shopping and Consumption Spaces." *Environment and Planning B: Urban Analytics and City Science* 48 (1): 132–50.

Geolytix. 2023. "Supermarket Retail Points." *Geolytix*. https://geolytix.com/blog/supermarket-retail-points/.

Graells-Garrido, Eduardo, Feliu Serra-Burriel, Francisco Rowe, Fernando M Cucchietti, and Patricio Reyes. 2021. "A City of Cities: Measuring How 15-Minutes Urban Accessibility Shapes Human Mobility in Barcelona." *PloS One* 16 (5): e0250080.

Green, Mark A, Konstantinos Daras, Alec Davies, Ben Barr, David Bayliss, and Alex Singleton. 2018. "Developing Openly Accessible Health Indicators for Small Areas in Great Britain: An Observational Study." *The Lancet* 392 (November): S39. https://doi.org/10.1016/s0140-6736(18)32876-9.

Haklay, Mordechai E. 2010. "How Good Is Volunteered Geographical Information? A Comparative Study of OpenStreetMap and Ordnance Survey Datasets." *Environment and Planning B: Planning and Design* 37 (4): 682–703. https://doi.org/10.1068/b35097.

Hobbs, Matt, Claire Griffiths, MA Green, A Christensen, and J McKenna. 2019. "Examining Longitudinal Associations Between the Recreational Physical Activity Environment, Change in Body Mass Index, and Obesity by Age in 8864 Yorkshire Health Study Participants." *Social Science & Medicine* 227: 76–83.

Jay, Jonathan, Felicia Heykoop, Linda Hwang, Alexa Courtepatte, Jorrit de Jong, and Michelle Kondo. 2022. "Use of Smartphone Mobility Data to Analyze City Park Visits During the COVID-19 Pandemic." *Landscape and Urban Planning* 228: 104554.

Mahabir, Ron, Anthony Stefanidis, Arie Croitoru, Andrew T Crooks, and Peggy Agouris. 2017. "Authoritative and Volunteered Geographical Information in a Developing Country: A Comparative Case Study of Road Datasets in Nairobi, Kenya." *ISPRS International Journal of Geo-Information* 6 (1): 24.

Overture Maps Foundation. 2023. "Overture Maps Foundation Releases Its First World-Wide Open Map Dataset – Overture Maps Foundation."

Owen, Danial, Daniel Arribas-Bel, and Francisco Rowe. 2023. "Tracking the Transit Divide: A Multilevel Modelling Approach of Urban Inequalities and Train Ridership Disparities in Chicago." *Sustainability* 15 (11): 8821.

Zhang, Liming, and Dieter Pfoser. 2019. "Using OpenStreetMap Point-of-Interest Data to Model Urban Change—A Feasibility Study." *PloS One* 14 (2): e0212606.