

Exploring Open and Crowdsourced Data for Transportation and Community Planning

A Research for the September 11th Memorial Program Fellowship
NYSDOT and NYMTC

Columbia University
EunJin Shin

Introduction



E.J. Shin

- 9-11 Memorial Fellow
- Columbia University
 - MS in Civil Engineering
 - Master of Architecture
- Co-founder at STUDIOIOIOIO
- AIA New York

Why Open/Crowd Source Data Now?

- Rise of “Big Data” in planning
- Growing public participation
- Real-time insight through geospatial tools



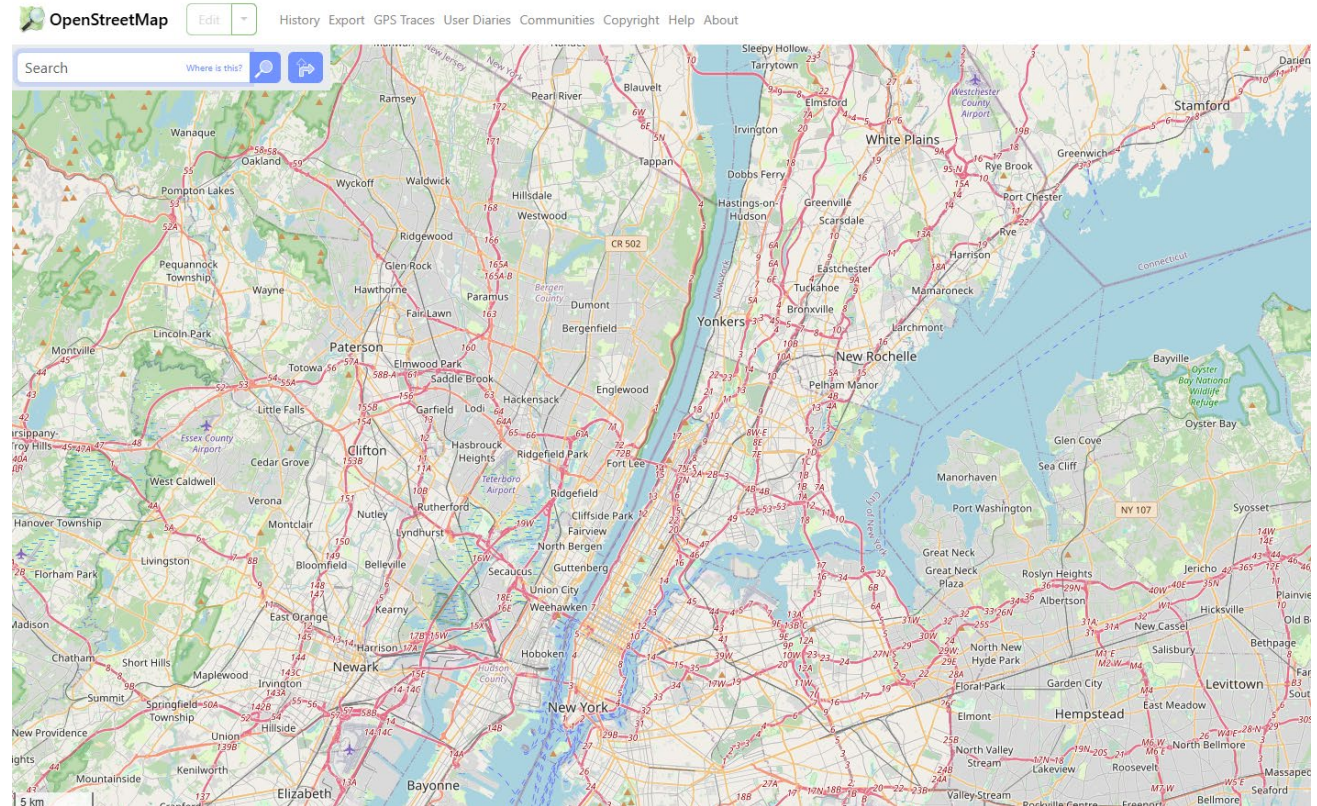
Research Objectives

Understand global best practices: five case studies

Identify scalable strategies for NYSDOT and NYMTC

What is OpenStreetMap (OSM)?

- User-generated geospatial database
- Used in planning, disaster response, infrastructure
- Funded by public-private partnerships



OpenStreetMap, www.openstreetmap.org. Accessed 22 Aug. 2025

Case Study Summary

	Transport for London	Land Transport Authority (LTA) Singapore	London Cycling Infrastructure Database (CID)	San Francisco Bike Lane	UrbanSim
Location	London, UK	Singapore	London, UK	San Francisco, CA	Global
Approach	“Data belongs to the public” engagement, trust	Real-time crowd monitoring through public Wi-Fi login	Delineation of cycling infrastructure	Public funded mobile application cycle tracks for data collection	NSF-funded project to develop planning simulation tool
Features	Real-time transport info for customers, optimizing bus services, managing traffic flow, growth planning	Passenger behavior insights on platforms (e.g. slower walking speeds) Can inform station design and operations	Direct OSM plug-in: infrastructure impacts on cyclist safety, cycling growth, economic effects of cycling	Public funded in-house open-source mobile app, data collection for in-depth analysis of bicycling behavior and infrastructure need	Open-source software, models urban land use, development, systems interactions over long periods. Used for policy development
Unique Aspects	Allows 3 rd party to develop software	Encryption algorithms utilized to protect privacy	High granularity & size, studies cycling economic impacts	Granular data collecting, developed by public funding	Microsimulation and comprehensive planning approaches

Case Study 1: Transport for London (TfL)

What it is: Public transport operator using open and crowdsourced data for real-time information and planning

Where: London, UK; ongoing since 2009 Open Data initiatives

Uses:

- Real-time travel info for staff and passengers
- Optimize bus services, traffic flow, city growth planning
- Public-reported maintenance and cycling/walking route planning

Data: Oyster cards, bus GPS, traffic lights, cameras; integrated with OSM and GTFS via open-source tools (OpenTripPlanner, GO_Sync)

GTFS = General Transit Feed Specification



<https://www.businessinsider.com/disappointing-photos-show-traveling-london-underground-really-like-2020-2>

Case Study 1: Transport for London (TfL)

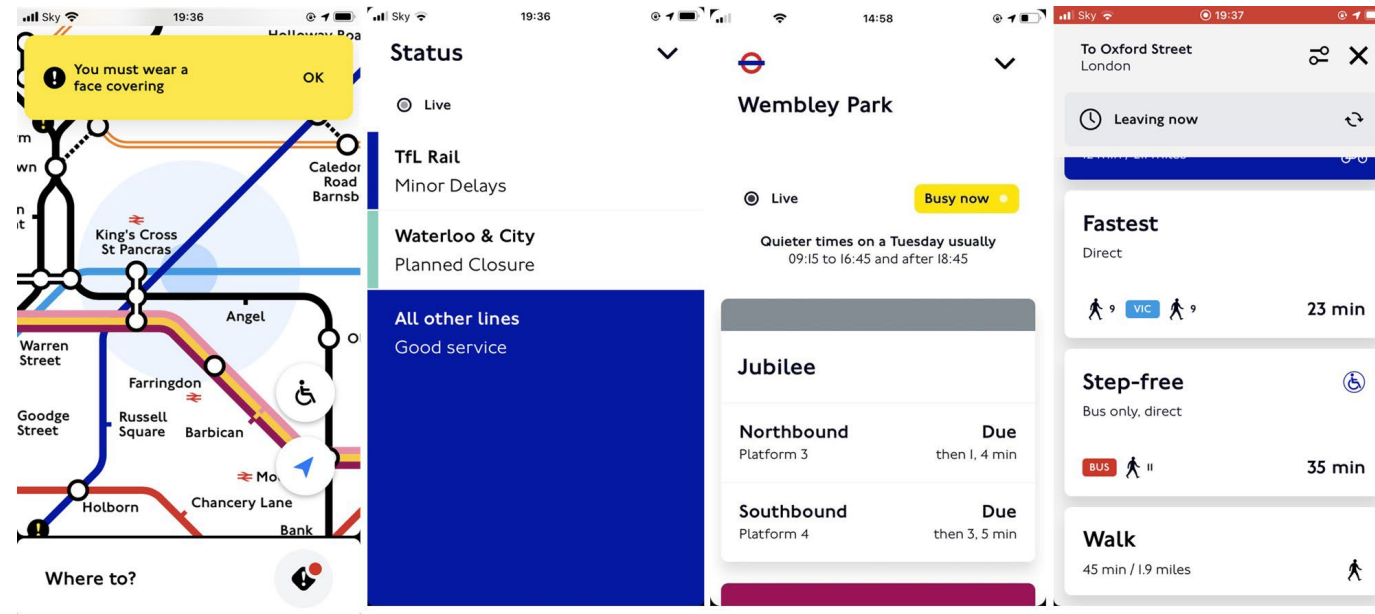
Impact

- Better service reliability
- Demand forecasting
- Efficient traffic management
- Higher public trust

Challenges

- Ensuring data quality
- Managing diverse data streams
- Presenting complex info clearly

Unique: Holistic integration of big data sources and strong 'data belongs to the public' philosophy; fosters innovation via third-party apps



“Transport for London Provides Real-Time Crowdedness Level on Tfl Go App.” *Mobility Innovators*, 1 July 2021, mobility-innovators.com/transport-for-london-provides-real-time-crowdedness-level-on-tfl-go-app

Case Study 2: Land Transport Authority (LTA) Singapore *Public Wi-Fi Crowd Heat Maps*

What it is: Public Wi-Fi logins in MRT stations, real-time crowd heat maps

Where: Singapore; 129k+ devices tested across Circle Line stations

Uses:

- Real-time crowd monitoring and prediction
- Train deployment and congestion management
- Passenger behavior and station design insights

Data: Wi-Fi sensors and big data framework; some OSM for risk maps and visualization



Cheung, M., Cheng, Y. & Fujiyama, T. Investigating passenger behaviour on the metro platform with Wi-Fi location tracking data: a case study of Singapore. *Transportation* (2024).
<https://doi.org/10.1007/s11116-024-10570-w>

Case Study 2: Land Transport Authority (LTA) Singapore *Public Wi-Fi Crowd Heat Maps*

Impact

- Improved service efficiency
- Accurate crowdedness prediction
- Passenger flow analysis

Challenges

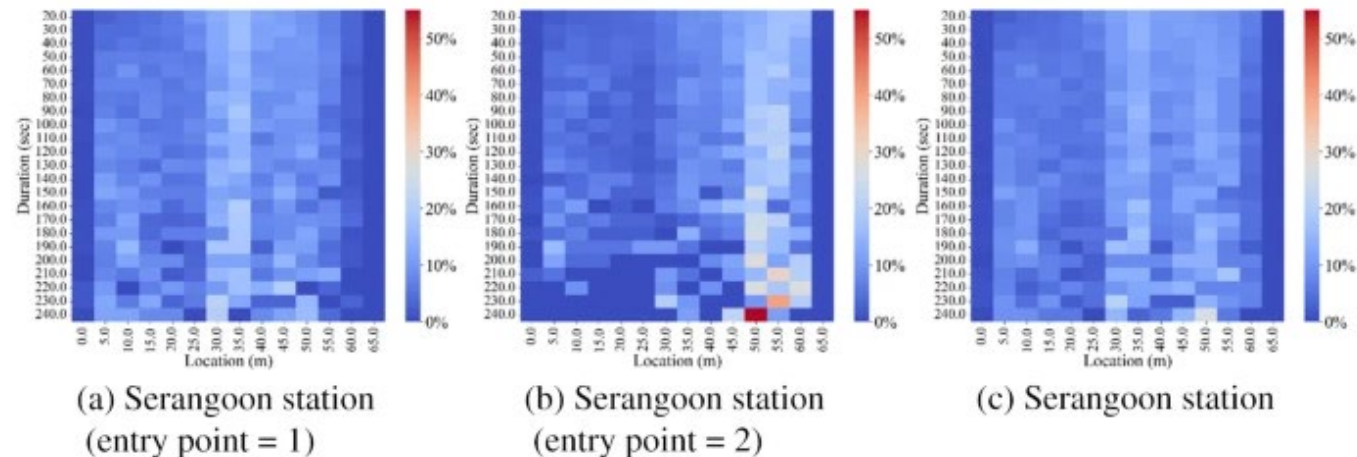
- Wi-Fi data requires cleaning
- Privacy managed via encryption
- Alternative to costly cameras

Unique: Non-invasive, cost-effective use of existing Wi-Fi infrastructure + ML (SMOTE + random forest) for crowd prediction

ML = Machine Learning

SMOTE = Synthetic Minority Over-sampling Technique

Random Forest = an ensemble machine learning for more accurate predictions



Cheung, M., Cheng, Y. & Fujiyama, T. Investigating passenger behaviour on the metro platform with Wi-Fi location tracking data: a case study of Singapore. *Transportation* (2024).
<https://doi.org/10.1007/s11116-024-10570-w>

Case Study 3 and 4: Bike Lane

London and San Francisco

Uses

- Inventory cycling facilities
- Assess safety
- Study economic impact on local businesses

Data Sources

OSM historical amenity data and crowdsourced inputs (danger spots, infrastructure quality)

- CycleTracks app (GPS)
- Strava Metro
- Bikeshare trip data
- OSM mapping



Case Study 3 and 4: Bike Lane

London and San Francisco

Impact

- Cycling growth and safer commutes (40 - 65% lower injury odds with protected lanes)
- \$3.3M investment → 14% cyclist increase
- 9,500 lane-block reports in nine months
- New shops and preference for bike lanes documented

Challenges

- Data quality and changing infrastructure mapping
- OSM contributor retention
- Participant recruitment
- Reliability of crowdsourced inputs

Unique

- First dataset at this scale linking cycling infra to business growth and economic activity
- Citizen reporting platform and tailored apps (CycleTracks) for municipal action

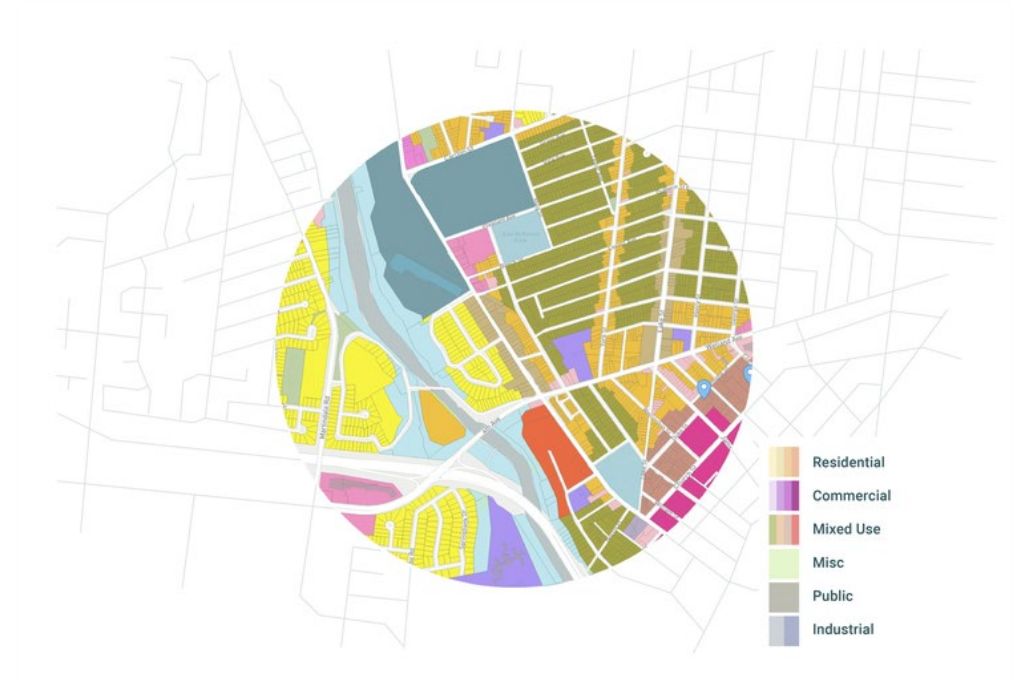
Case Study 5: UrbanSim

- **What it is:** Open-source tool to model land use and transportation over 20+ years
- **Where:** NSF-funded (USA), applied in US, Europe, Taipei
- **Uses:** Test policy impacts (transit, zoning, incentives) on growth, transport, energy, CO₂
- **Data:** Open-source, integrates with OSM; outputs as maps/charts
- **Impact:** Neutral, evidence-based support for planning decisions
- **Challenge:** Needs detailed data; OSM gaps
- **Unique:** Microsimulation (households/businesses), integrates land use and transportation

Case Study 5: UrbanSim



3D Rendering



2D Mapping

“UrbanSim Artificial Intelligence for Sustainable Development with AI.”
UrbanSim, www.urbansim.com Accessed 22 Aug. 2025

Building a Roadmap

Hybrid data models

Institutional partnerships

Privacy-focused, equity-driven strategies

What Could Go Wrong?

- Data quality and contributor fatigue (due to volunteer base work)
- Gaps in coverage (rural, underserved areas)
- Privacy and consent



Key Takeaways for NYSDOT and NYMTC

Promote public ownership of data

Integrate diverse datasets (e.g., GPS, fare cards, civic apps)

Democratize planning with open-source tools

Conclusion and Vision

- **Toward Open-Source Urbanism**
- Empowering citizens as co-creators – potential marketing and leverages
- Planning that is dynamic, inclusive, and evidence-informed
- A roadmap for resilient, responsive systems



Special Thanks to My Wonderful Advisors!!

- Dr. Ali Afshar
 - Ellwood Hanrahan
 - Celeste Harp
 - Jim Davis
 - Richard Batchelder
 - David Rosenberg
-
- Seitu Allen
 - Joomee Lee



Thank you!