# Capturing Urban Mobility using Crowdsourced Data from Smartphones

Dr. Yang Wang and Dr. David Philip McArthur

Urban Big Data Centre

An ESRC Data Investment

# Introduction

- Project funded by InnovateUK

- The aim is to encourage innovation and new commercial ventures

- Our project is to produce detailed mobility data

# Some current data sources and limitations

- Surveys
  - Small samples
  - Expensive
  - Hard to complete
  - May lack detail

- Sensor networks
  - Requires (expensive) infrastructure
  - Lacks origin/destination info
  - Not multimodal

Urban Big Data Centre

catch

# (Some) Big data
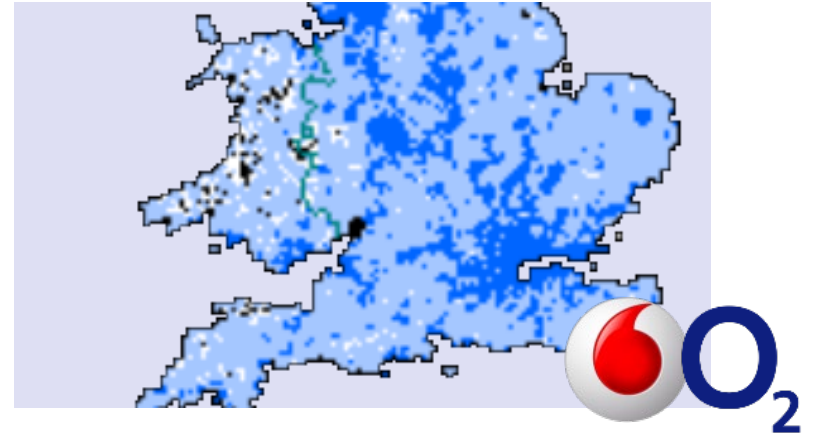
# Some limitations

- Hard/expensive to access

- Pre-processing of data can be a black box

- May be tied to one mode

- Spatial coverage may be patchy

Urban Big Data Centre

catch

# Smartphone data

- Smartphones offer the chance to collect rich data
  - Independent of operator
  - Information for all travel modes
  - Full door-to-door OD and route data
  - Potentially real-time

# Catch! Project

- The Catch! (citizens at the city's heart!) app is a journey planning app

- It passively collects GPS trajectories

- The phones' sensors and machine learning are used to infer travel mode

- Users get to contribute data to improve transport planning in their city

- Insight from the data can feed back into better journey suggestions

# The Consortium

## SMEs

- TravelAi, The Behaviourlist, Elgin Roadworks, Placr
- *App development, data sources, citizen messaging, impact assessment*

## Research organisations

- University of Glasgow (UBDC), University of Leeds (CDRC), Transport Systems Catapult
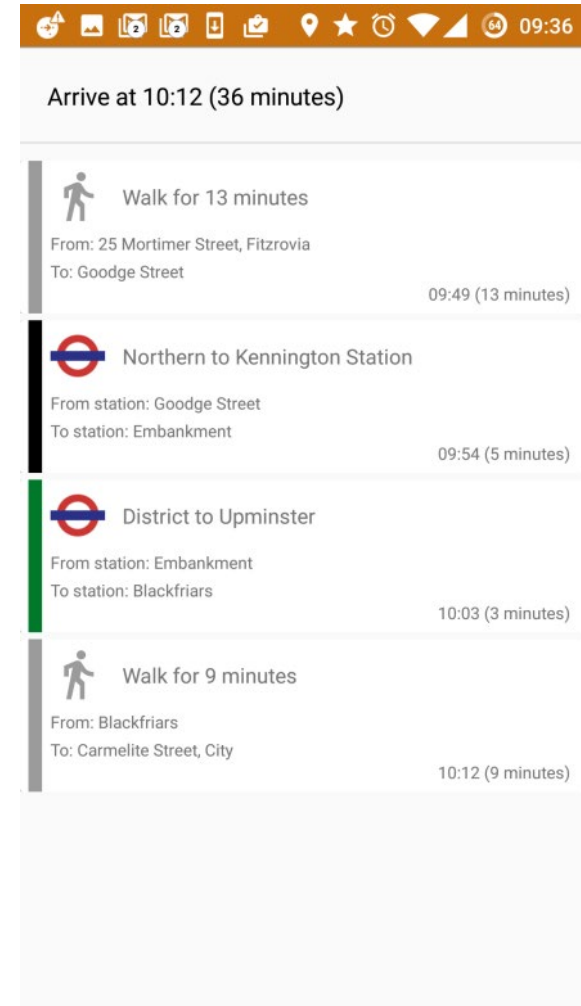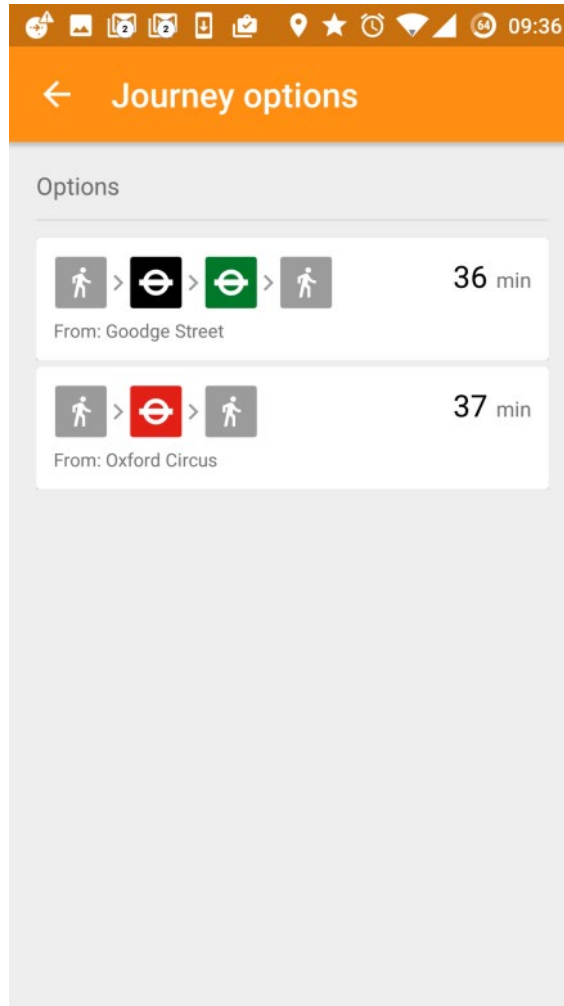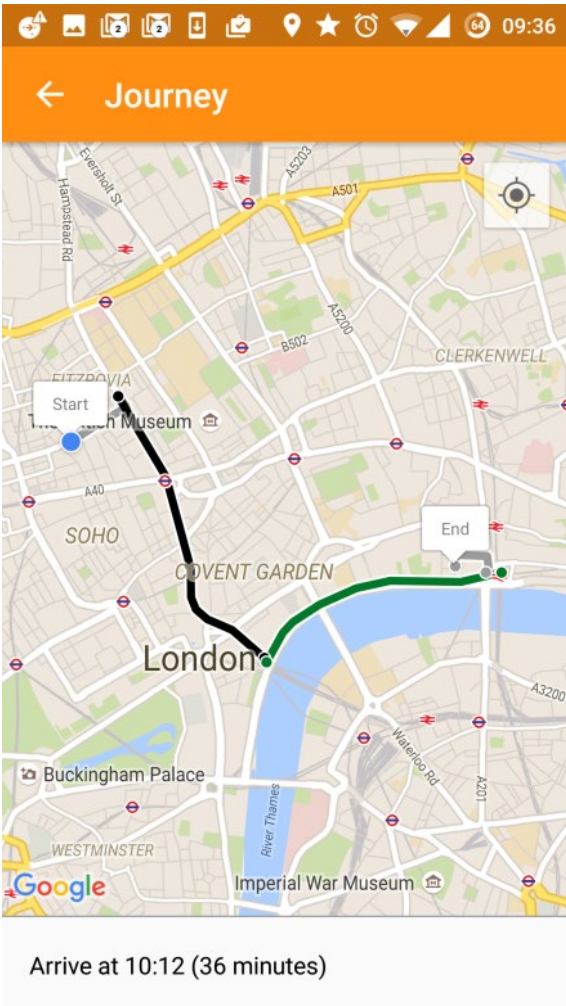- *Data cleaning, anonymization, aggregation, analysis*

## Local Authorities and cities

- Coventry, Ipswich, Leeds, Newcastle, Oxfordshire
- *Citizen access, sounding boards, pilots, data sources, advocates*
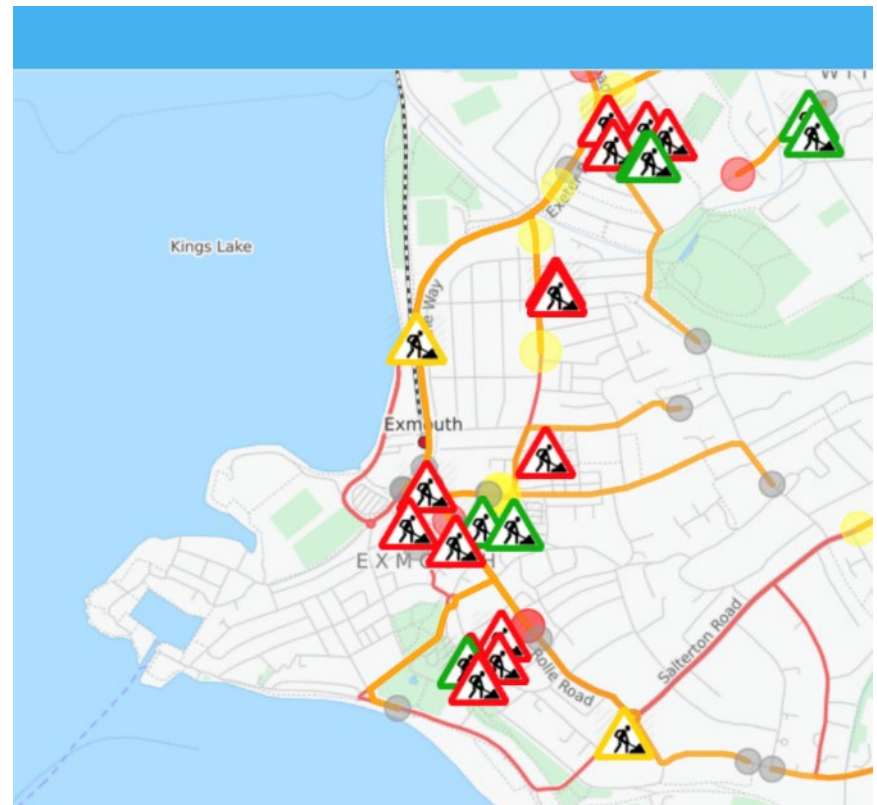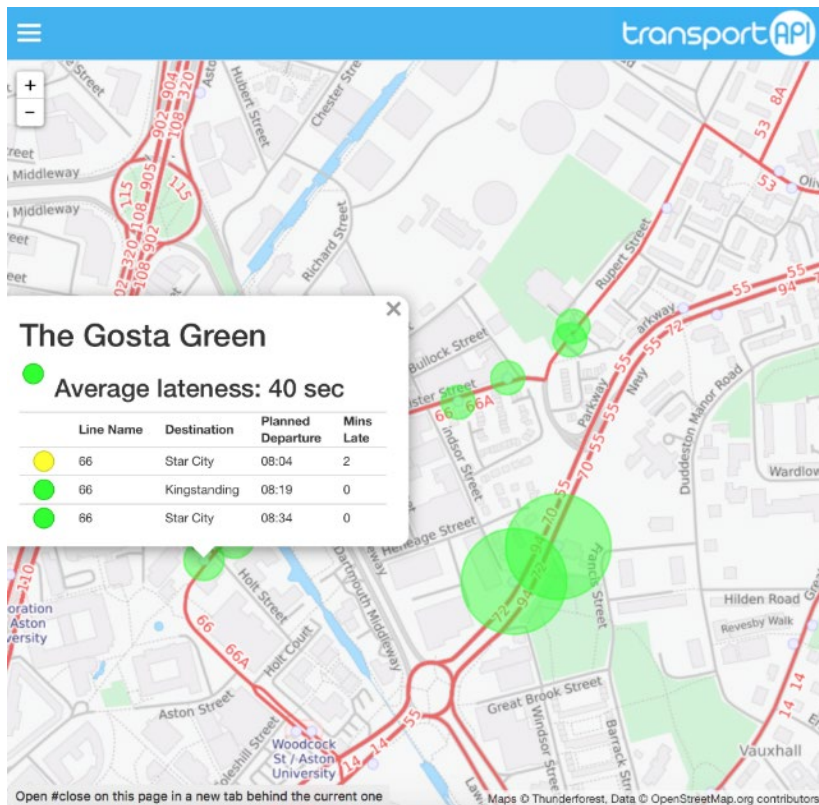
Urban Big Data Centre

catch

# The App

# Catch! App

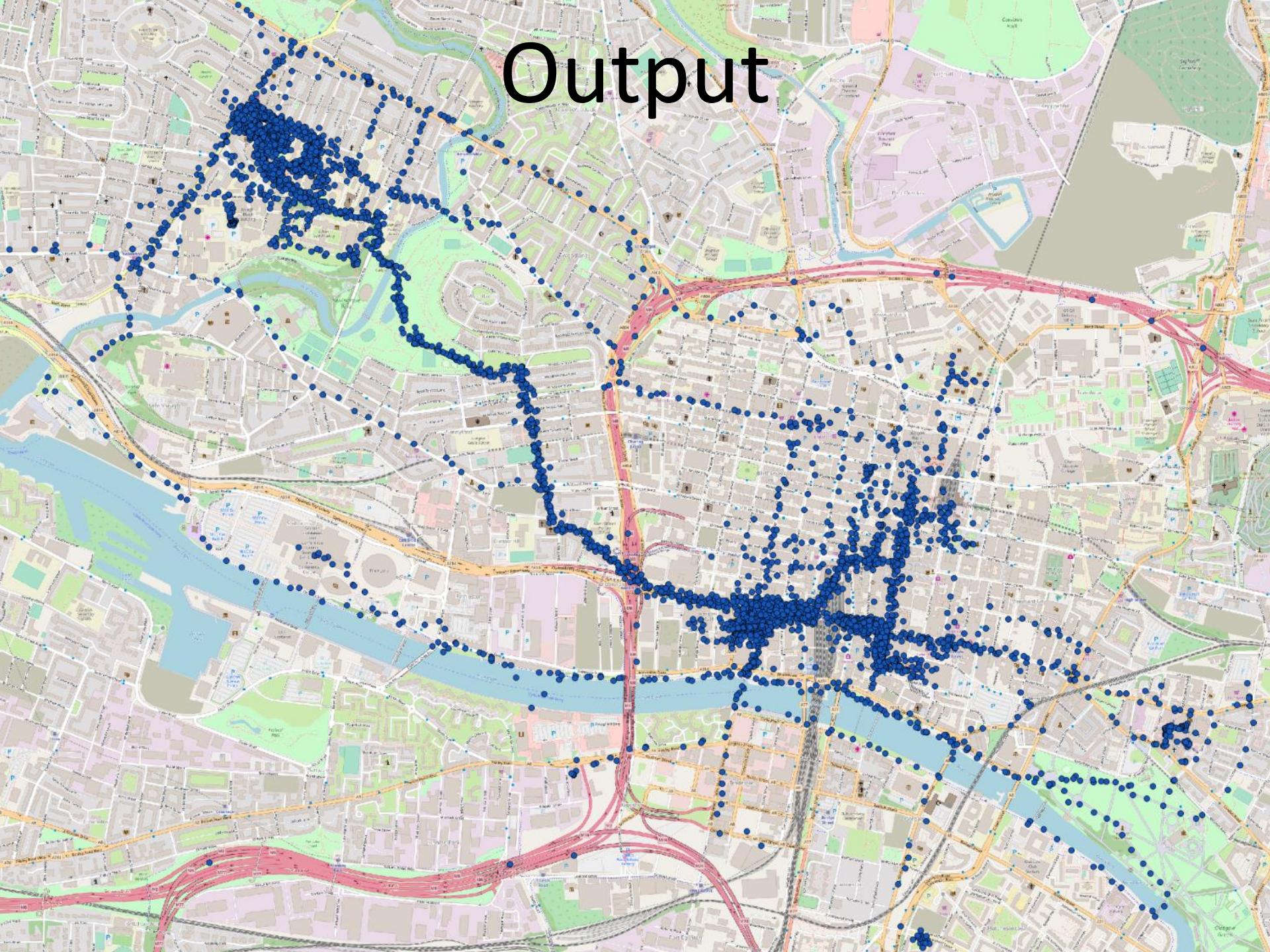- Includes real time information on roadworks (Elgin) and public transport performance (TransportAPI)

# Data from the app

- Person identifier

- Latitude, longitude

- Time

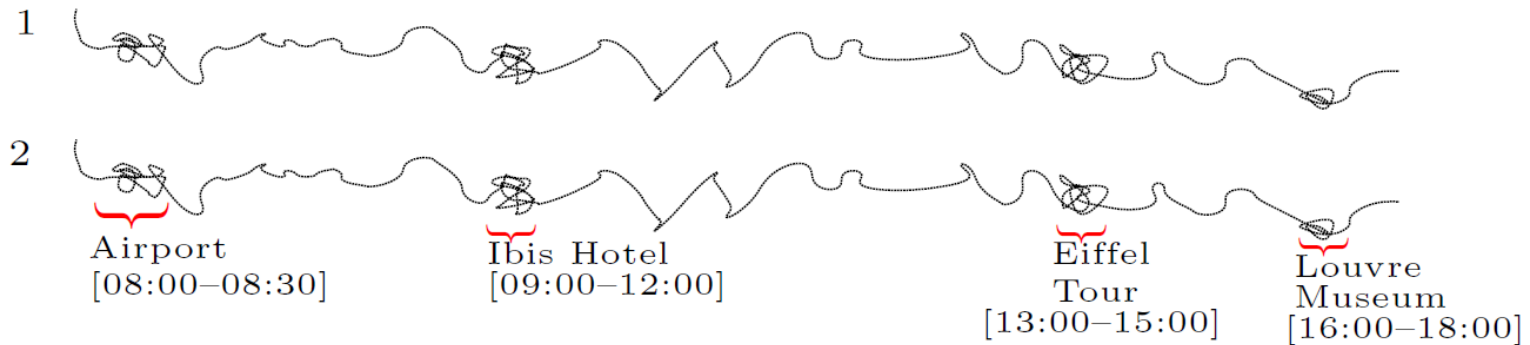- Inferred mode

- Collected every 5 seconds
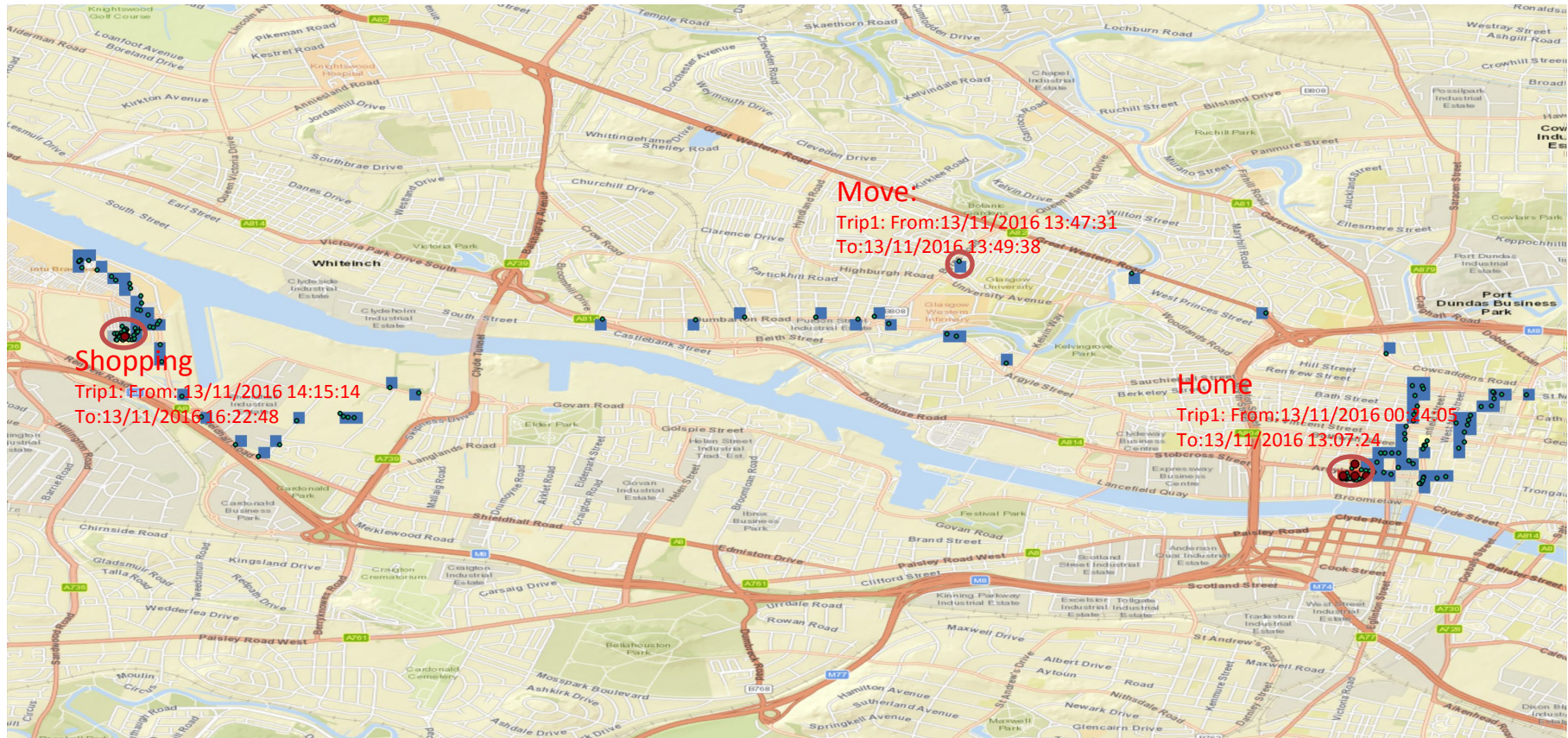
# Our part of the project

- Clean and enhance the data
  - Map match the data
  - Detect and annotate stops
  - Aggregate the data into a useful, privacy-protecting format
  - Improve mode detection with post-processing

# Semantic Trajectory



Alvares, L.O., Bogorny, V., Kuijpers, B., Macedo, J.A., Moelans, B. and Vaisman, A. (2007a), "A model for enriching trajectories with semantic geographical information", Proceedings of the 15th ACM International Symposium on Advances in Geographic Information Systems, ACM, Seattle, WA, pp. 1-8.
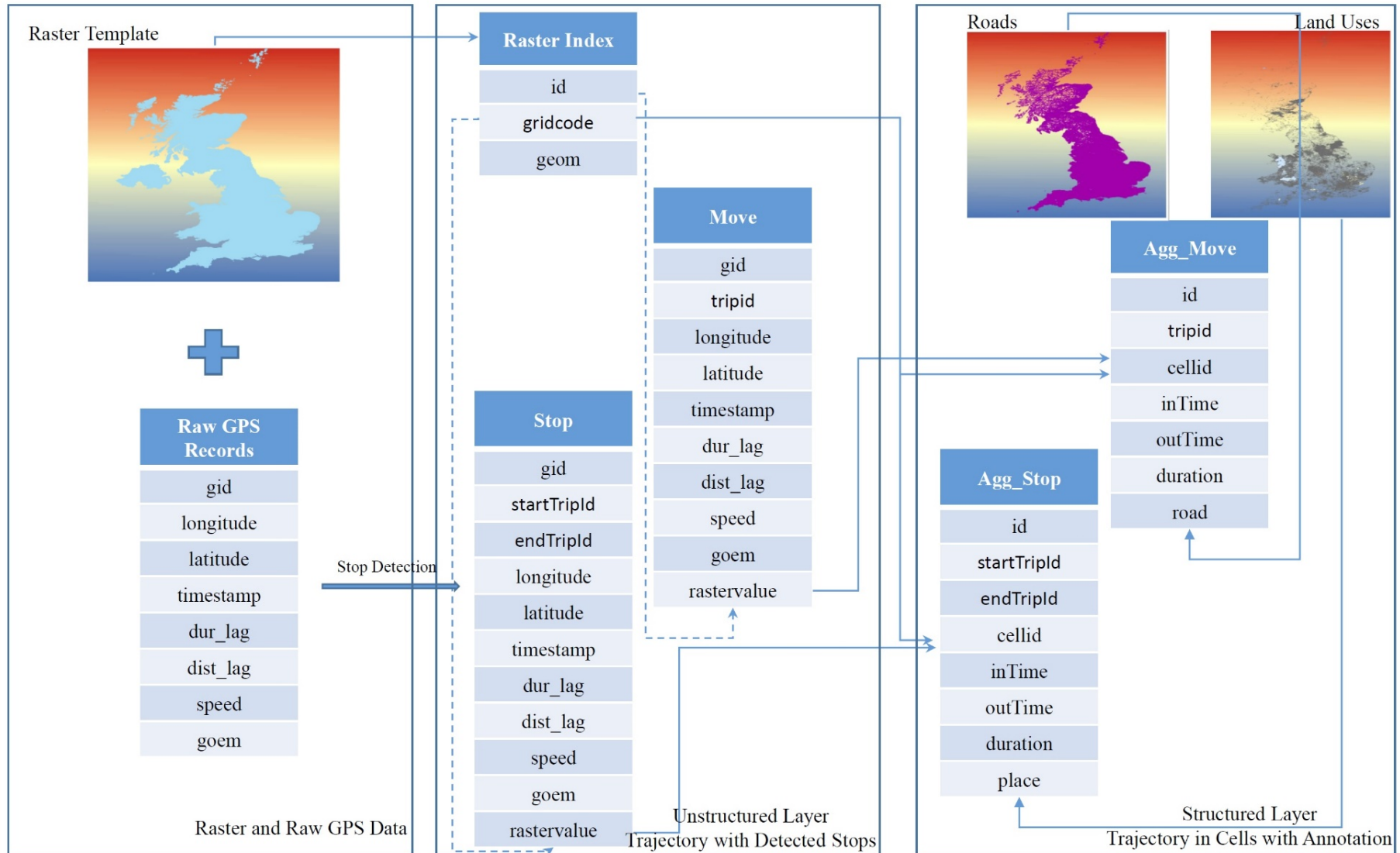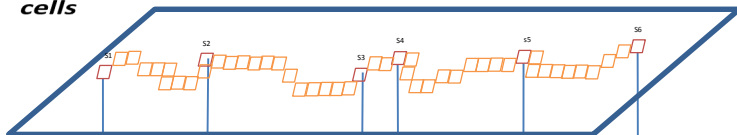
# An Illustration of the Method
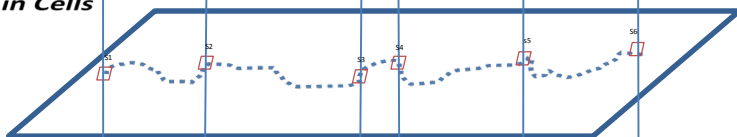
# Semantic Trajectory Framework
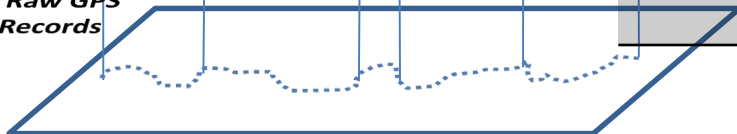
# Stop/Move Table Structure



**Structured trajectory in cells**

**Spatio-temporal Trajectory with Stops in Cells**

**Raw GPS Records**

## Move

| Id | tripId | CellId | inTime | outTime | Duration | *Road* |
|---|---|---|---|---|---|---|
|  |  |  |  |  |  |  |

## Stop

| Id | startTripId | endTripId | CellId | inTime | outTime | Duration | *Place* |
|---|---|---|---|---|---|---|---|
|  |  |  |  |  |  |  |  |

## Move

| Id | tripId | ts | Duration_lag | Longitude | Latitude | Distance_lag | Geometry | CellId |
|---|---|---|---|---|---|---|---|---|
|  |  |  |  |  |  |  |  |  |

## Stop

| Id | startTripId | endTripId | ts | Duration_lag | Longitude | Latitude | Distance_lag | Geometry | CellId |
|---|---|---|---|---|---|---|---|---|---|
|  |  |  |  |  |  |  |  |  |  |

## Raw

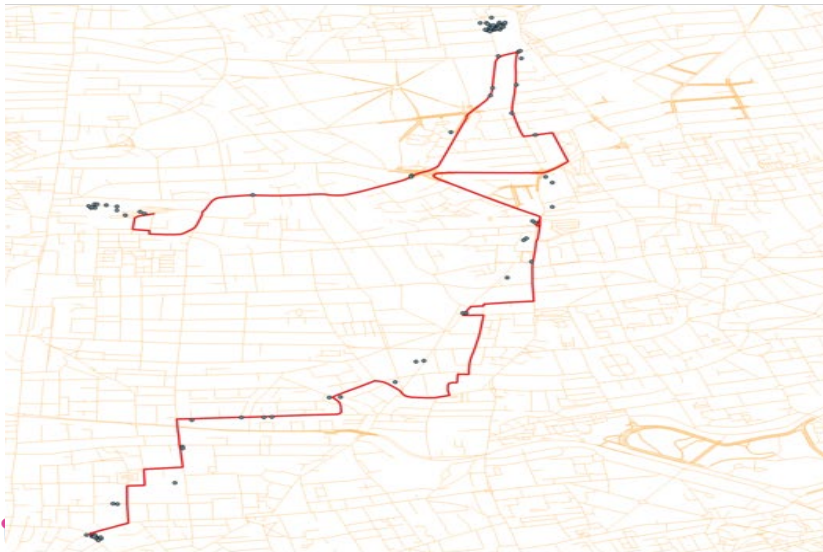| Id | ts | Duration_lag | Longitude | Latitude | Distance_lag | Geometry |
|---|---|---|---|---|---|---|
|  |  |  |  |  |  |  |

Urban Big Data Centre

catch

# Travel Mode Detection
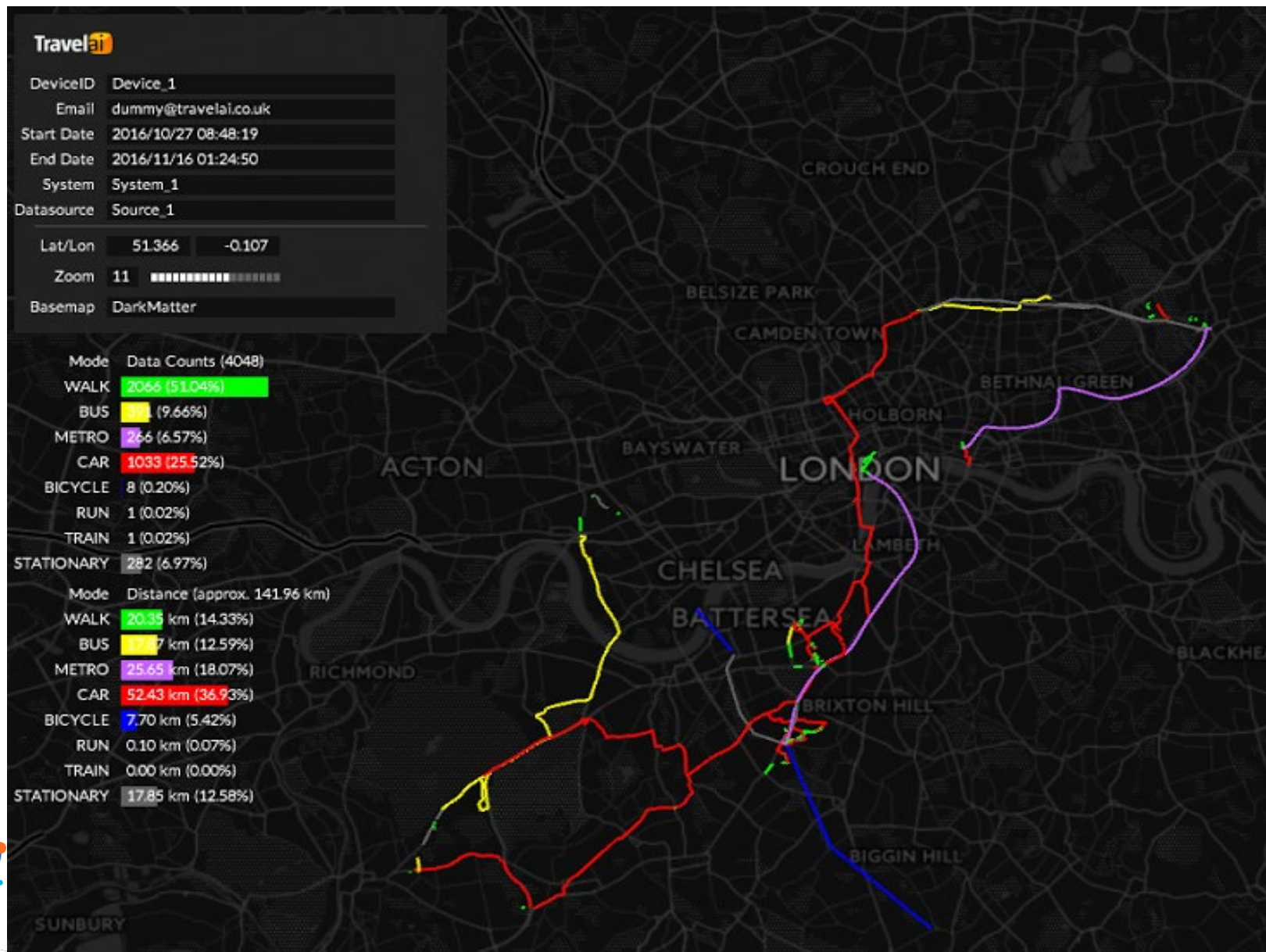
# Map Matching to Road Network



mm

MHT

GraphHoppers
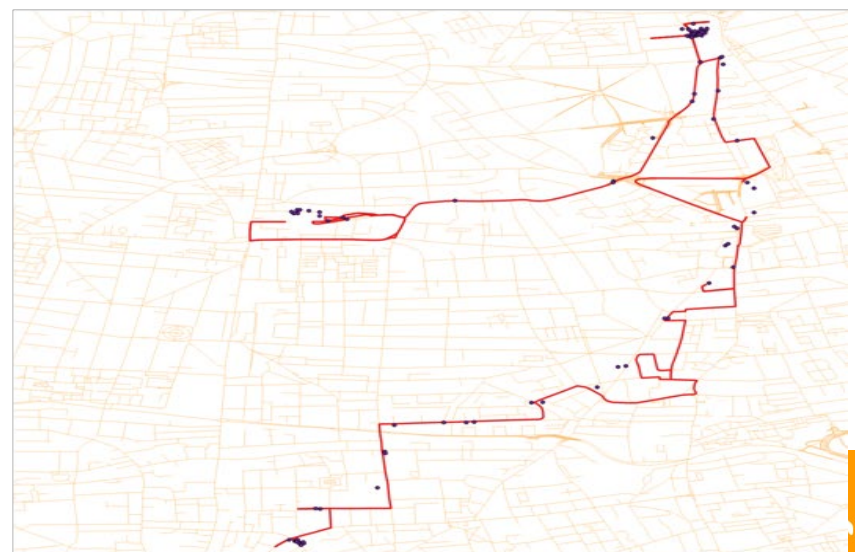
Barefoot

# Semantic Annotation
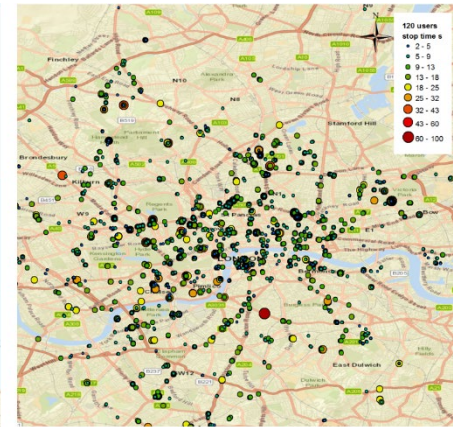


Moves with roads

Example of Annotated Trip

Move:
Trip1: From:13/11/2016 13:47:31
To:13/11/2016 13:49:38

Shopping
Trip1: From: 13/11/2016 14:15:14
To:13/11/2016 16:22:48

Home
Trip1: From:13/11/2016 00:34:05
To:13/11/2016 13:07:24
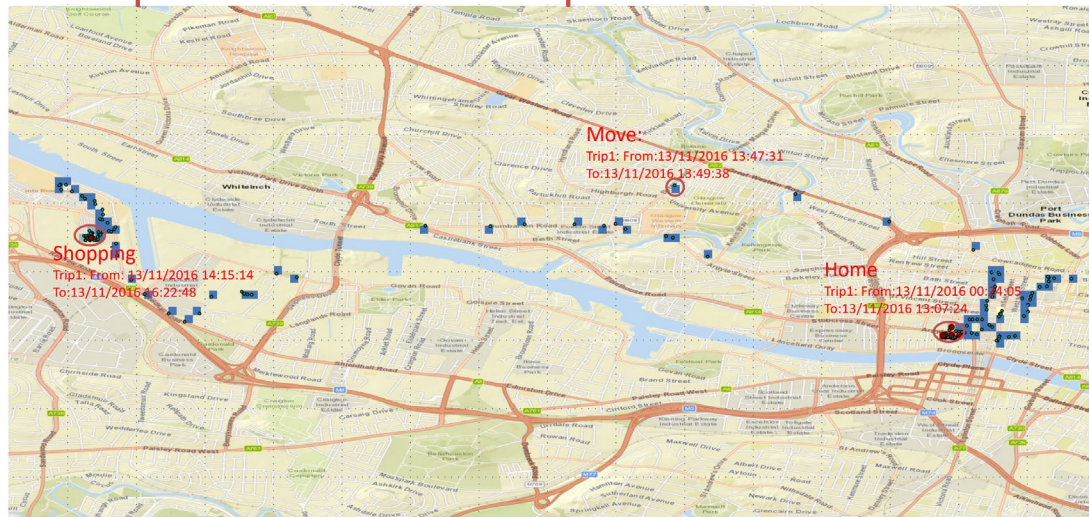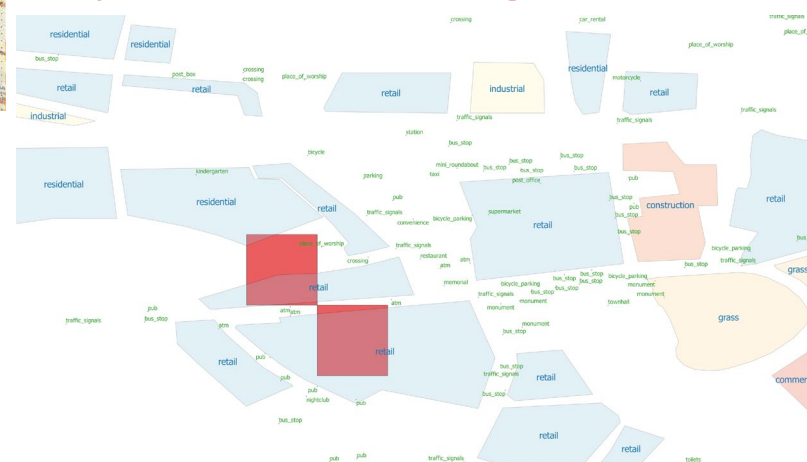
Aggregated Travel Speed on Roads

Aggregated Stop Time on Road Intersections

Stops with surrounding information

Urban Big Data Centre

# Privacy Issue – A Real Case

## Focus on two schools

1. App adopted by pupils and parents
2. Data analysis:

   Quantification of transport modes

   Correlation with geography

   Correlation with congestion

   Correlation with air quality

   Identification of related journeys

3. Long term benefits

   Reduce congestion and improve air quality

   Shift to healthier and eco-friendly transport

   Input into school travel plans

   Improved town planning

   Car sharing and child safety apps

   Better transport routing

   Base data for traffic modelling

Urban Big Data Centre

catch

# Anonymization - Sharing Aggregated 'Moves'

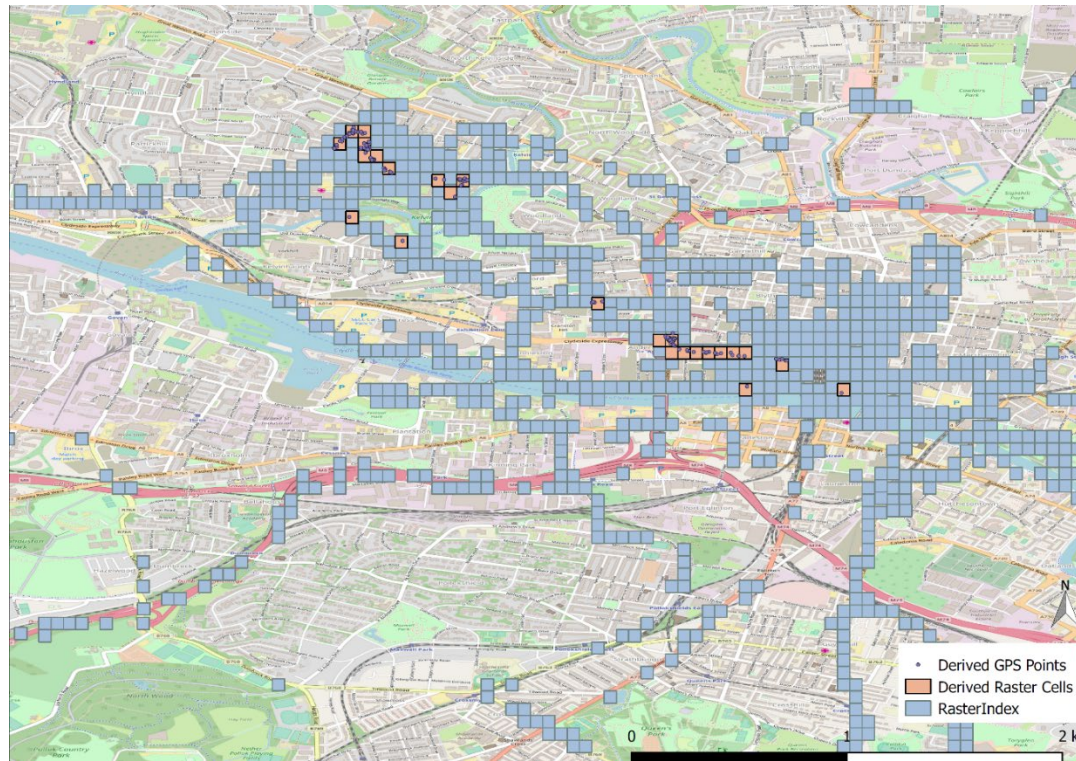- Trip count within temporal window
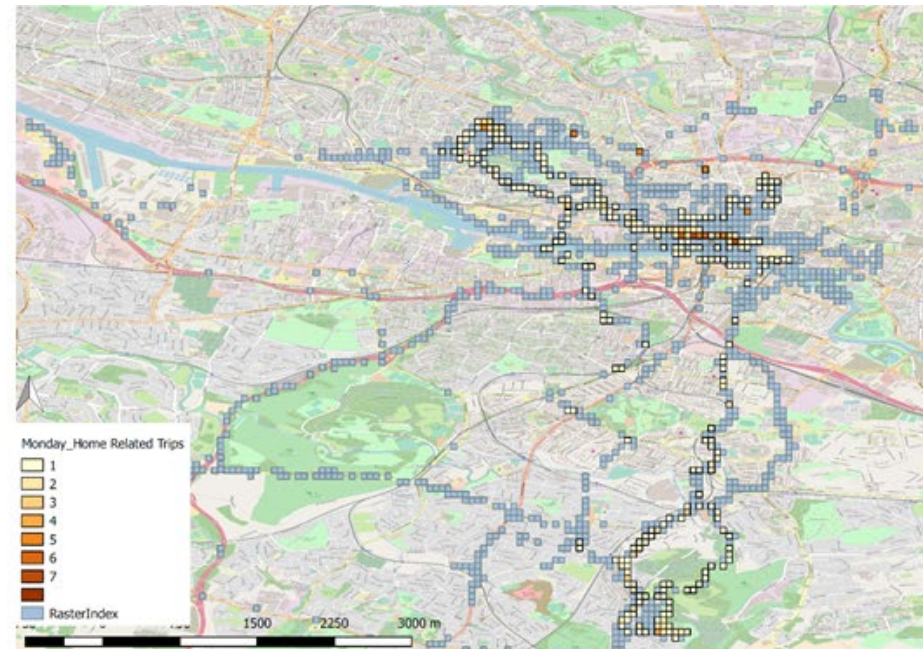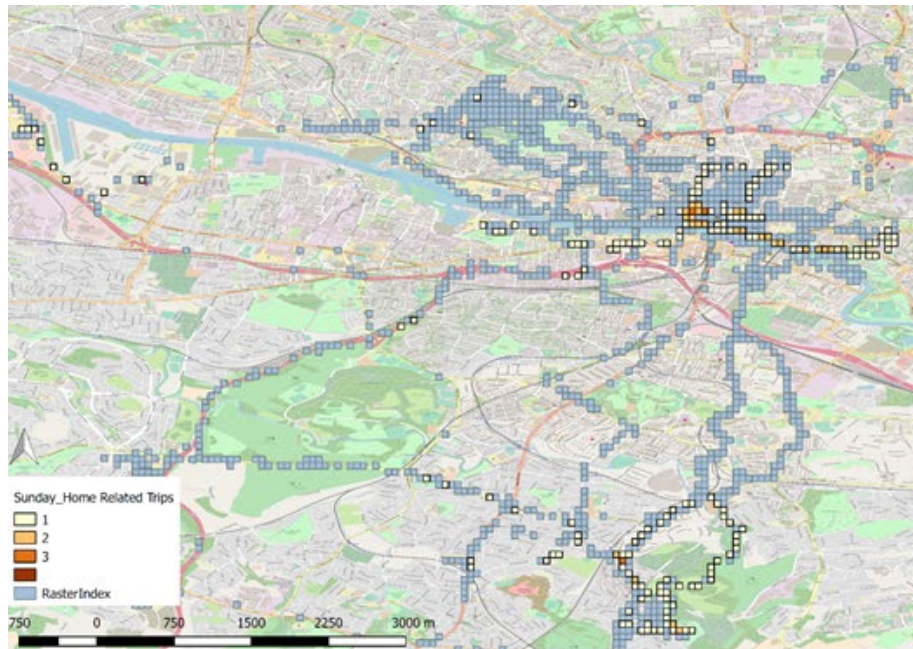


Example of extracting home related trip counts for Glasgow area on Sundays (Left) and Mondays (Right) symbolized in trip counts against the whole raster coverage (RasterIndex) in blue
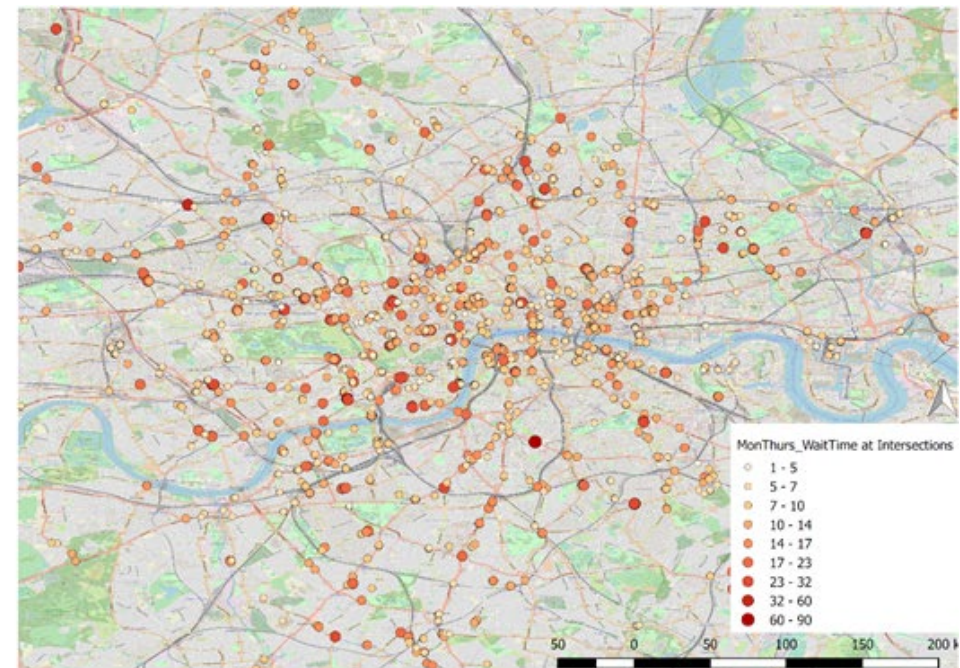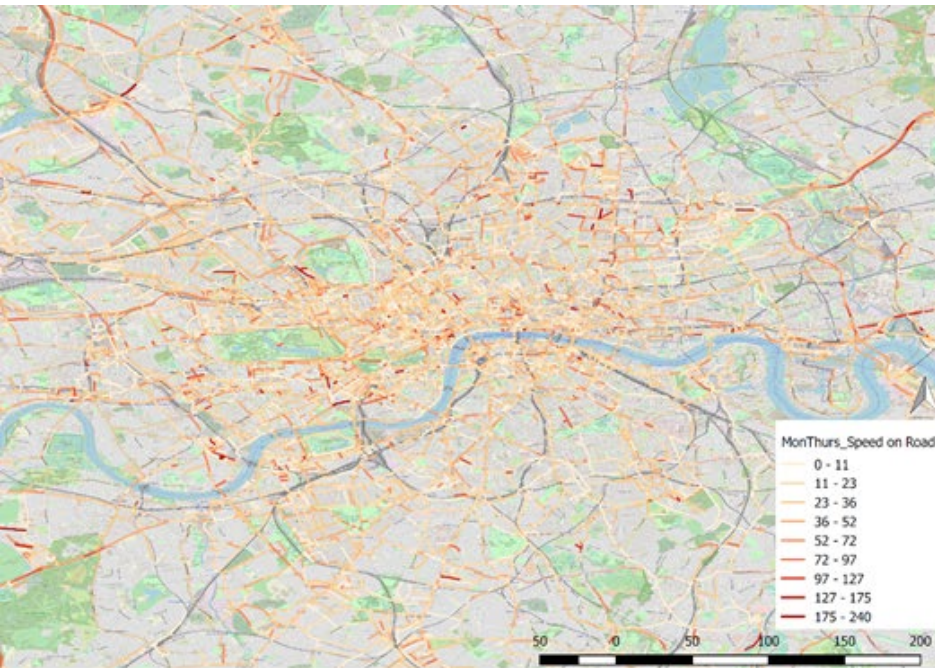
# Anonymization - Sharing Aggregated 'Moves' -2

- $k$-anonymity within temporal window



Shared 'move' cells confirming to $2$-anonymity in Glasgow within 17:00-18:00-time window for use.

# Anonymization - Sharing Aggregated 'Moves' -3

- Aggregation of GPS attributes on roads within temporal window
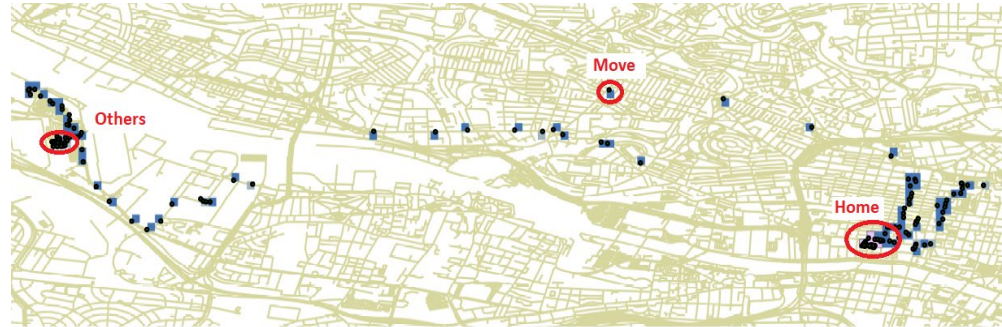


Aggregated GPS information on roads for 120 users.
Left is aggregated travel mean speed (km/h) on road segments symbolized from yellow-red.
Right is aggregated wait time (a) on road intersections.

# Anonymization - Sensitive Locations - 1
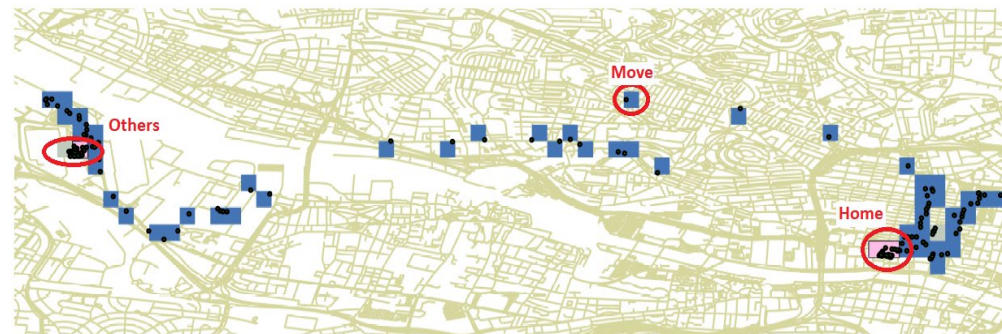
- Grid-masking

  Grid-masking result with different cell sizes



Single Sized Raster Cell
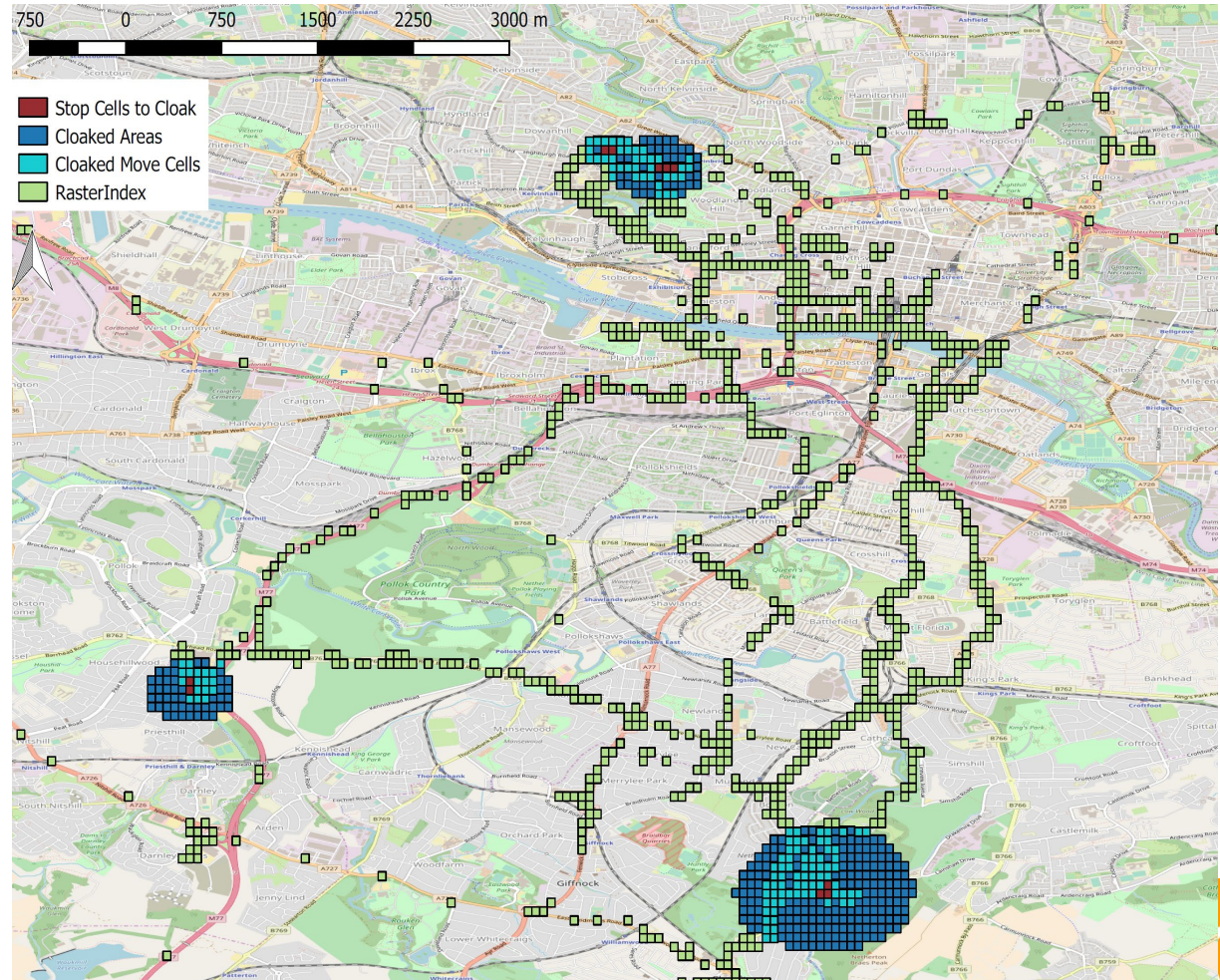


One and a Half Sized Raster Cell



Double Sized Raster Cell
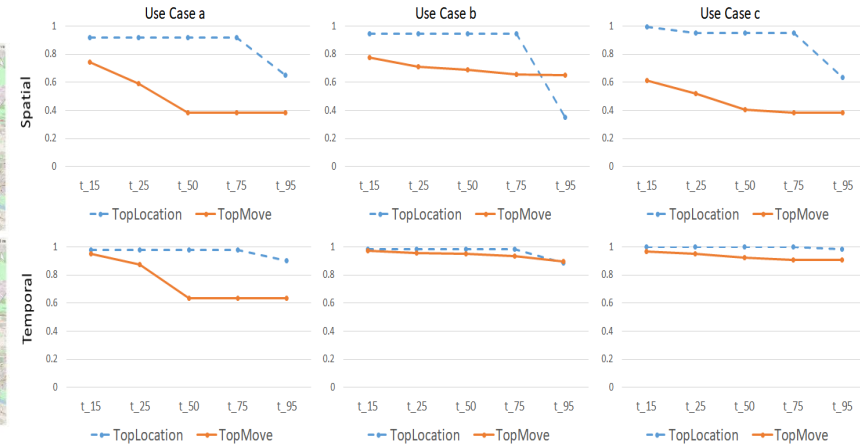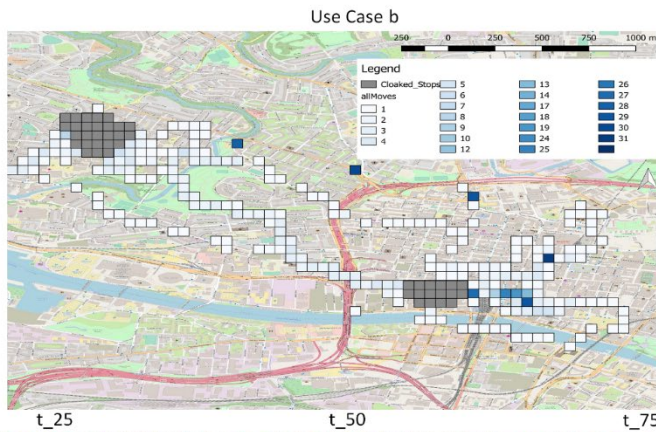
# Anonymization - Cloaking Sensitive 'Stops'

- Blurring stops into the environment

Blur the stops in the 'home/work' category into the 10 nearest buildings and the stops in the 'other' category into the 10 nearest points of interest.

# Anonymization - Cloaking sensitive 'moves'



- Zoomed into specific purpose of trips;
- Zoomed into temporal window;
- Information loss for
  - selection of methods;
  - selection of thresholds.

# (Some) limitations of the project

- Will the app be used by enough people?

- The data are biased; but how? Will this change over time?

# To download the app

- Not always easy to find in the App/Play Store due to the name

iOS

Android

catch

# Thank you for your attention.

Wang, Y., & McArthur, D.P. (2018). Enhancing data privacy with semantic trajectories: A raster-based framework for GPS stop/move management. *Transactions in GIS*.

@UrbanBigData

@UofGlasgow