Integrated Multimedia City Data (iMCD) Project

Urban Big Data Centre

Mark Livingston, iMCD programme manager & Katarzyna Sila-Nowicka, Research Fellow
Urban Big data Centre

• Bringing together and linking data from wide variety of sources
  – Local government and business

• Providing new and innovative data sets to
  – Academics;
  – Policy makers
  – and general public
iMCD collaborators

• Urban Studies
  – Transport

• Computer science
  – 2 information retrieval teams

• Education
  – Adult learning team
iMCD data strands

- Household survey (incl. every adult)
- GPS and Lifelogging
- Satellite data/ LiDAR
- Multi-media data retrieval
- Textual media data retrieval
iMCD Survey

• Household survey
• Target 1500 households
• Expected between 1,710 to 2,850 people (60 to 100%)
• Coverage was Glasgow Clyde Valley planning area (all contiguous LAs to Glasgow + Inverclyde)
iMCD Survey

• Main modules on:
  – Travel and Transport
  – Sustainability
  – ICT
  – Education and Learning

• Collecting data on Attitudes, behaviours and Knowledge
iMCD Survey

• Demographics
• Neighbourhood and communities
• Previous residence
• Health
• Employment
• Income
• Cultural and civic engagement
Example questions

SUPPOSE YOU HAD £100 IN A SAVINGS ACCOUNT AND THE INTEREST RATE WAS 2\% PER YEAR. AFTER 5 YEARS, HOW MUCH DO YOU THINK YOU WOULD HAVE IN THE ACCOUNT IF YOU LEFT THE MONEY TO GROW? WOULD IT BE ...

More than £102 [1]
Exactly £102 [2]
Less than £102 [3]
I can not tell, not even approximately [4]

DOES YOUR HOUSEHOLD HAVE ANY PET DOGS?
Yes/No
iMCD

• In field 15\textsuperscript{th} April 2015 to 21\textsuperscript{st} November 2015.
• 1508 households accepted(51%)
• 30% households sampled refused
• 2095 people too part (74% of eligible adults)
• 1.4 adults per household
• Compared to 1.9 adults per household in census
• Data at LA level should be available very soon.
GPS and Life-logging devices

trajectory point
(Lon, Lat- WGS 84)

GPS device

Autographer

Urban Big Data Centre
An ESRC Data Investment

University of Glasgow
Approximately what proportion of our survey respondents do you think took part in the sensor survey?
### Device selection

Table 5.2: Comparison of GPS loggers; Five models of trackers were evaluated in 12 categories and marked from 1-4, where 1 means weak and 4 - strong performance of a tested tracker.

<table>
<thead>
<tr>
<th>Category</th>
<th>GlobalSat DG-100</th>
<th>Atmel BTT308</th>
<th>Transmit 747 ProS</th>
<th>I-got U 600</th>
<th>QSTARZ 1000XT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Battery life</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Functions</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Data recorded</td>
<td>1</td>
<td>3</td>
<td>4</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>Settings</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Charging</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Carrying/wearing</td>
<td>1</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>Safety issues</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Ease of use</td>
<td>4</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Data format</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>Ease of download</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Accuracy of the data</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Price</td>
<td>4</td>
<td>2</td>
<td>4</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>Overall</td>
<td>26</td>
<td>30</td>
<td>37</td>
<td>27</td>
<td>35</td>
</tr>
</tbody>
</table>
# Activity Diary

## Example

📅 If the places are home, workplace or school, you do not need to write the Nearest Intersection Streets.

### 1. Commuting trip:

Assume that you (alone) drive to the park and ride and take a subway to get to your workplace. The transport modes you use are: Car -> walk (to the station from the parking lot) -> subway -> walk (to your workplace). This trip starts from 8:00am to 8:45am.

<table>
<thead>
<tr>
<th>Travel method</th>
<th>1st</th>
<th>2nd</th>
<th>3rd</th>
<th>4th</th>
<th>5th</th>
<th>Name of place &amp; activity (from LIST 2)</th>
<th>Nearest Intersection Streets</th>
<th>Start</th>
<th>End</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vehicle make/model/year</td>
<td>Honda Civic 2007</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TIME in minutes</td>
<td>10</td>
<td>5</td>
<td>23</td>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### 2. Shopping trip:

Assume that you take a subway to get to the St. Enoch shopping mall in the centre. The transport modes you use are: Walk (to the station from the parking lot) -> subway -> walk (to the shopping mall). This trip starts from 2:00pm to 2:32pm.

<table>
<thead>
<tr>
<th>Travel method</th>
<th>1st</th>
<th>2nd</th>
<th>3rd</th>
<th>4th</th>
<th>5th</th>
<th>Name of place &amp; activity (from LIST 2)</th>
<th>Nearest Intersection Streets</th>
<th>Start</th>
<th>End</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vehicle make/model/year</td>
<td>John Lewis &amp; Buchanan st &amp; South St</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TIME in minutes</td>
<td>10</td>
<td>20</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

---

| Participant | Date | Trip ID | Mode1 | Mode2 | Mode3 | Mode4 | Mode5 | Place | Activity | Address | Start_time | End_time | Vehicle_mode | Vehicle_year | duration_in_minutes | duration_Mode1 | duration_Mode2 | duration_Mode3 | duration_Mode4 | duration_Mode5 |
|-------------|------|--------|-------|-------|-------|-------|-------|-------|---------|---------|------------|----------|--------------|---------------|-----------------|---------------|--------------|---------------|--------------|---------------|--------------|
| 1070201     | 27/05/2019 | 1 | 1 | - | - | - | - | Kilmacolm | 12 & 10 Torr Road | 11:20 | 12:10 | Ford S Max | - | - | - | - | - | - | - |
| 1745401     | 28/05/2019 | 1 | 1 | - | - | - | - | Fraser St | 12 & 10 Argyll & Buchanan St | 10:35 | 12:30 | - | - | - | - | - | - | - |
| 1745402     | 28/05/2019 | 2 | 1 | - | - | - | - | Card Shop, Hyndland | 7 & 10 Hyndland & Highburn | 14:00 | 14:34 | Volkswagen Pl | 2011 | 3 | 30 | 30 | 30 | 30 | 30 | 30 |
| 1745403     | 28/05/2019 | 3 | 1 | - | - | - | - | Sainsbury's | 12 & 10 Hyndland Rd & Dumbarton Rd | 20:20 | 20:30 | Volkswagen Pl | 2011 | 3 | 30 | 30 | 30 | 30 | 30 | 30 |
| 1745404     | 28/05/2019 | 2 | 1 | - | - | - | - | 5 | - | - | - | - | - | - | - | - | - | - | - | - | - | - |
| 1745405     | 28/05/2019 | 1 | 1 | - | - | - | - | 5 | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - |
| 1745406     | 28/05/2019 | 2 | 1 | - | - | - | - | Home | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - |
| 1051601     | 23/05/2015 | 1 | 3 | 3 | 3 | 3 | 3 | Town Centre & The Mall | 6 & 10 | 19:45 | 15:20 | - | - | 65 | 10 | 10 | 10 | 10 | 10 | 15 | 20 |
| 1051602     | 23/05/2015 | 2 | 3 | 3 | 3 | 3 | 3 | Port Glasgow | 9 & 10 | 19:20 | 16:20 | - | - | 60 | 10 | 10 | 10 | 10 | 10 | - | - |
| 1051603     | 23/05/2015 | 3 | 3 | 8 | 8 | 8 | 8 | High Street, Port Glasgow | 9 & 10 | 19:05 | 09:05 | - | - | 50 | 10 | 10 | 15 | 15 | 15 | - | - |
Designed systems
### Statistics

<table>
<thead>
<tr>
<th>LL + GPS</th>
<th>Valid data from number of respondents</th>
<th>Invalid data from number of respondents</th>
<th>Average Age for Invalid data respondents</th>
<th>Average Age for Valid data respondents</th>
<th>Number of respondents</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female</td>
<td>125</td>
<td>22</td>
<td>36.36</td>
<td>47.78</td>
<td>147</td>
</tr>
<tr>
<td>Male</td>
<td>102</td>
<td>16</td>
<td>51.06</td>
<td>46.92</td>
<td>118</td>
</tr>
<tr>
<td>sum</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>265</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>GPS</th>
<th>Valid data from number of respondents</th>
<th>Invalid data from number of respondents</th>
<th>Average Age for Invalid data respondents</th>
<th>Average Age for Valid data respondents</th>
<th>Number of respondents</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female</td>
<td>67</td>
<td>9</td>
<td>37.00</td>
<td>46.24</td>
<td>76</td>
</tr>
<tr>
<td>Male</td>
<td>54</td>
<td>8</td>
<td>45.50</td>
<td>49.64</td>
<td>62</td>
</tr>
<tr>
<td>sum</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>138</td>
</tr>
</tbody>
</table>
Input
Data processing

Trajectory segmentation

Stops Identification

Semantic Development

Output

GPS data

Cleaned Data \([x,y,t]\)

Raw Trajectories

Structured and Segmented trajectories

Semantic trajectory

Based on Spaccapietra, 2002
Image analysis
Which techniques can be used to protect privacy with GPS data?
Textual media data and multimedia data retrieval
Twitter

- Tweets **geolocated** in Glasgow: we use a **polygon around Glasgow** (-4.3932, 55.7953; -4.0913, 55.9212)

- Tweets from **certain users** e.g. @BBCWestScot, @policescotland

- Tweets from with **certain terms or hashtags**: e.g. glasgow, #glasgow2014

- **Format**: JSON contains text of posts and meta-data (e.g. user, time, location if any, etc.)

- **Quantity**: 65 million tweets, 4.1 Terabytes (compressed)

- **Duration**: July 2014 – November 2015
What percentage of tweets in our sample are geo-tagged?
Environmental Data

Weather

- **WorldWeatherOnline.com**: Glasgow’s weather (hourly weather information)

- **Format**: JSON (information about temperature, humidity, precipitation, wind direction, pressure etc.)

- **Duration**: Collected between December 2014–November 2015

- **Quantity**: 9615, hourly data-points
Transportation Data

Trains and Cycle routes

<table>
<thead>
<tr>
<th>Static</th>
<th>Real time</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Cycle routes</strong> – created by data team In Glasgow City Council. Specified the streets/roads that have cycle routes</td>
<td><strong>Rail delays</strong>: a daily data feed of the timings of passengers trains (from Network Rail via <a href="http://realtimetrains.co.uk">http://realtimetrains.co.uk</a>)</td>
</tr>
<tr>
<td><strong>Quantity</strong>: 416 roads</td>
<td><strong>Quantity</strong>: 8 million train services</td>
</tr>
<tr>
<td><strong>Railway stop names</strong>, locations, identifiers (NaPTAN dataset)</td>
<td></td>
</tr>
<tr>
<td><strong>Quantity</strong>: 2602 stops</td>
<td></td>
</tr>
</tbody>
</table>

GPS Traces

- Latitude, longitude, speed, altitude of 300 survey respondents
Urban Data Dashboard

Unified portal for all services and data. Restricted by IP address as per UBDC ToS.

Social Media Exploration Services

Data Services

Social Media
Explore search, browse and navigate the social media content we are collecting through our portal. Provide access to the content that we retrieve from this content, by exploring visualisation, networks, and text processing techniques that are tailored to social media context.

Transportation
Explore transportation mode from via blogged GPS data in the city. Urban transportation information about traffic delays in the city.

Weather
Explore search, browse and navigate the social media content we are collecting through our portal. Provide access to the content that we retrieve from this content, by exploring visualisation, networks, and text processing techniques that are tailored to social media context.

Tweet Terms or Hashtags
Query the tweet collection based on a hashtag or specific tweet terms.

Events in Twitter
Examine the events detected by our automated systems related to specific terms of interest.
Transportation Data Exploration

Delayed train timeline querying
Obtain a timeline of delays for a given train station at a given time

Visualise GPS traces and transport modes
Obtain a map of the GPS traces and obtain a breakdown of the transport modes
Real time train delays feed

- Train delays service (timeline) for 1st July 2015:

Train delays peak on the 1st July 2015
Tweet data search

• Search tweet stream for terms indicating negative sentiment regarding train transport:

Commuter unhappiness also peaks on the 1st July 2015
Tweeting Behaviour during Train Delays

At 8am it was already **26C in West London and 20C in Glasgow!**....take a closer look at your location. #heatwave http://t.c…"

@Ms_Hilz: UK can't handle anything....snow, problem....**mid 30's heat, they are calling it "wrong heat"** and to prepare for train cancella…"

@ScotRail can't cope with the weather when it's hot or cold apparently #cancellations

When your train has lost signal in this heat and **everyone is trapped on the train boiling**

Due to the heat there are speed restrictions on our trains and some **cancellations**' Come on, the train on Spirited Away

People are passing out because of the heat on this train. No air whatsoever in the carriages, which are completely packed
Thank you for your attention

Questions ?