(Mobile network data
6-month acceleration project)

A datathon for the public by the public[1]

Table of Contents

---

[1] Project name should be named with communications support and involvement of the operators!

1.0 Background and objectives

The Digital Economy Act empowers ONS and the Government Statistical Service (GSS) to work much more effectively and authoritatively to access private sector data (Appendix 1: *Understanding what the Act means for ONS commercial approach to accessing private sector data*). Following some early ONS progress with mobile network data (looking at origin to destination commuting), and mindful of wider public sector success (transport related use cases have been tried and proven in major government departments and devolved administrations) ONS (with a lead from Data Science Campus and the Data as a service team in Methods, Data and Research) is looking to better engage the UK's mobile network operators ███████████████████████ to assess how we can accelerate the potential of mobile network data to better inform national policy making. We kicked off a roundtable discussion with the three primary operators at the Data Science Campus launch and have since conducted 1:1 interviews with each. Each operator demonstrate a different approach to collecting and processing mobile network data and a different breadth of experience developing public interest use cases. This project and approach will enable ONS to better tap into that expertise, develop the right relationships and secure the data access needed in line with our new opportunity: The Digital Economy Act.

Our immediate goal (next 6 months) is to showcase how these innovative data can support the statistical community to unlock new capability and insights for all of us. Into the medium term (12 months+) we would hope to have firmed up more frequent use of these data to augment or potentially even replace existing statistical approaches and have a clearer view on how best to process and access the data in perpetuity. This project will balance the strategic need for ONS to improve the timeliness and authority of outputs with the imperative of driving down data collection costs. Several overarching questions will be validated in our approach:

> ➢ Can mobile network data be used to reproduce, augment or create entirely new existing statistical outputs?
>> o How does the quality and timeliness compare to traditional approaches?
> ➢ How reusable are these data sources – can they be frequently combined with other data (i.e. survey or administrative)?
> ➢ Is there an opportunity for ONS to co-create a methodological approach with the operators for using these data?
>> o Can we develop a core data standard and or core dataset working with the operators (see 3.3)?
> ➢ Can available flows of mobile network data support ONS to produce valuable outputs faster and for less?
>> o Can we quantify these potential savings / improvements?

2.0 Data privacy, confidentiality and communications

Public perception is a critical consideration in this work. A communications led approach to ONS, operator and external stakeholders will be taken from the start and we shall resource independent external data privacy expertise to ensure a rigorous approach to personal data privacy. It is generally accepted that accessing anonymised individual level data versus aggregated counts has the potential to support broader and more in depth research (e.g. a wider range of mobility applications) but the trade off is greater ethical and privacy implications. This project will further explore these challenges and look at solutions to overcoming them (e.g. taking methodology/algorithms *to* the data instead of accessing the data could be one such approach).

3.0 Project approach

An initial six-month project (May - October 2017) goal will focus on applying mobile network data to tackle a shortlist (probably 2/3) of prequalified policy questions as well as to lay a foundation to better exploit the potential value in securing meaningful longer-term data access. The policy themes below will be distilled down by the ONS team. GSS Heads of Profession's input has also been fielded – the Health and Safety Executive may be an appropriate partner for one or more of the themes and we are working with them to better validate this.

Toward the end of the six months project ONS will run a 'Datathon' style event to bring together public and private sector teams of data scientists and engineers applying collaborative intensive effort in a fixed duration research '*Datathon'* style event (likely 2-3 days in Sep - Oct 2017). Ideally the recently launched Data Science Campus will host this event. The format of the event will be structured to focus the very best data science methods and expertise on pre-prepared datasets against carefully curated policy 'Challenges.' These Challenges will be curated and run in an approach similar to the Nesta and ODI Open Data Challenges series.[2] In the build up to the Datathon event the Office wants and needs to work openly and collaboratively with government, the mobile network operators and associated data science community to ensure the most pressing and relevant policy questions for these data are developed and that datasets curated will be, in principle, capable of answering them. We cannot expect to answer every Challenge we define but we very much want and believe this approach we will uncover new insights and learn how best to make use of mobile network data for the longer term. We also plan for the engagement to help ONS forge the right relationships with operators – ideally leveraging the deep, somewhat varying use-case domain experience they each have. We will use and develop existing ONS research and capability in this area and ensure we are mindful of and learn from wider international work in this area (e.g. ESSnet and CBS Netherlands have pockets of expertise).[3]

---

[2] http://www.nesta.org.uk/project/open-data-challenge-series
[3] NB reference to Big Data team papers: Mobile phone data use across other NSI's and the GSS

3.1 Policy challenge sector themes

Two/three data Challenges and associated questions must be distilled from the following range of policy themes and use cases below and developed with the operators. Appendix 3 (Statistical Applications for Mobile Network Data) provides a more detailed list and external references. The following list is a proposed shortlist from which ONS leadership need to further shortlist two/three priority areas.

3.1.1 Urban planning and land use

International studies show the relative densities of Call Detail Records (CDR's), either across different areas or across time (NB. these methods do not develop population estimates.

As well as having application directly within the production of land use statistics (such as business, residential, retail, nightlife etc), intelligence on where and at what locations people are to be found in greater numbers may be of use within operational processes such as conducting survey fieldwork. By extension, the use of this type of information in close to real time can provide valuable intelligence for other operational uses such as the optimisation of crowd control procedures and other services such as site location for ambulances etc.

3.1.2 Mobility

Understanding how and where people move to can help to target services to areas of greatest need. For this reason mobility analysis has been used extensively in urban planning and transport applications. Studies have considered the relative densities of CDR's, either across different areas or across time (as for urban planning) although anonymised individual-level data is considered to be better for mobility applications. Research using a time series of passive mobile phone positioning data at individual level can identify locations that are repeatedly visited by mobile phone users. These meaningful locations may be inferred as "home" and "work". By aggregating individual-level data for which home and work have previously been derived, it is possible to produce home-work origin-destination flows. Such flows have been compared, at different spatial scales (as low as 100x100metres in dense urban centres). Various outputs such as distance to work, time to commute, # of employed (who commute to a regular place of work) (all of interest to Census/Labour Force Survey). Example use cases: UK wide journey to work / commuter patterns, UK wide travel choices (e.g.http://mobile.abc.net.au/news/2017-03-27/sydney-squeeze-sydney-needs-to-become-a-smart-city/8389452?pfmredir=sm)

3.1.3 Tourism

A detailed study by Eurostat assesses the feasibility of using mobile positioning data within inbound, outbound and domestic tourism statistics from several aspects including access, cost, trust, and the technological and methodological challenges. Tourism indicators include: number of trips or visits; number of nights spent; number of days spent and; number of unique visitors. These indicators may be broken down further by: country of residence; aggregations by time (day, week, month); aggregations by geography; duration of trip or stay (same-day or overnight trip); main destination; secondary destination, transit pass-through;

collective movement patterns and repeat visits. There is also research investigating the counts of international country codes of every call or SMS made and received by mobile phone users (in Milan, Italy) with a spatial resolution of about 200 metres. The researchers show that the observed spatial distribution of international codes well matches the distribution of international community's reported by official statistics. They also investigate robust clustering patterns that can be used to identify the touristic hotspots.

3.1.4 Ethnicity

An approach to assigning ethnicity is to use the language setting on the mobile phone as a proxy indicator, research in Estonia and in South Asia shows this. Both these studies declare a high correlation is found between the language setting and underlying ethnicity. The studies then go on to examine the mobility behaviour of ethnic minority groups and find differences in segregation to the majority ethnic group and also to census data that is based on residential population rather than day-time population. If shown to be capable of being replicated in the UK, these research approaches might be used in similar official statistics applications such as in the identification of some ethnic minority populations and how integrated these populations are with the population as a whole. However, inferring measures of identity such as nationality, religion and ethnicity is highly contentious and raises strong ethical issues.

3.1.5 Socio-economic status and economic levels

CDR's have been used to identify the socio-economic status and economic levels in a population. This might be useful as covariate information for use in small-area modelling within official statistics of income. Most of the research uses data from developing countries where reliable census or survey data is sparse. Variables such as the number of calls made and received, the average distance covered whilst making calls and the distance calls are made to or received from have been shown to be related to an area's socio-economic level with good accuracy. Mobile phone network analysis has also been done to make the observation that more insular communities are likely to be poorer (i.e. Richer areas are more likely to communicate with each other often.)

3.1.6 Population and migration

Daytime population (DTP) can be summarised as the number of people per municipality, postal code, grid cell, etc. over time and can be seen as special case (aggregation) of mobility, it is much easier to produce (ONS was really close in 2015), and it allows publication on higher regional detail. Such an output also underpins other applications such as commuting so consideration to the overlapping utility of this use case should be given. ESSnet and CBS Netherlands have also done work in this area.

Population by time of day can be also be reported by applying methods to estimate population density from call or subscriber volumes using a variety of data formats from aggregated and individual level call detail records (CDR). These methods tend to identify a relationship between call/connection volumes during the night and official estimates of residential population, the population basis for most official estimates. Generally, once a reliable relationship is found, it is extrapolated to produce population estimates directly from

call/connection volumes at different times of the day/season. It has also been shown that this relationship may be used to downscale a known national population total to smaller areas. Policy questions might include:

- How might mobile network data provide an alternative approach to the Annual Population Survey and or augment/improve it?
- Can ONS access network cell counts over time to understand better the movement of people over time and in different locations?
- Can these data support/augment or replace insights developed from the UK/Passenger surveys?

3.1.7 Industrial strategy

How might mobile data better assist decision making relating to questions in the government's recently published industrial strategy? Questions in this category will span other sector headings listed: mobility, economic growth, workplace patterns and needs etc.

3.1.8 Emergency services and disaster / incident response

Mobile network data has been leveraged in instances where natural disasters have occurred (e.g. Haiti, New Zealand) but less has been done to better understand how systems and services in cities might react and respond (one emergency services exception is: http://london-fire.labs.theodi.org/). The Health and Safety Executive GIS and statistical head of profession has expressed an interest in collaborating with ONS regarding workplace incident management as are keen to better understand the movement of people in densely populated urban areas before and after workplace incidents.

- Can these data better prepare emergency responses teams on our streets?
- Can these data be combined with other sources in this category to develop new insights? (e.g. crime prediction https://motherboard.vice.com/en_us/article/mobile-phone-data-can-help-predict-crime-a-month-in-advance

3.2   Cost benefit analysis

The newly formed business development function in ONS Finance led by Simon Duddy can provide a unique capability to provide cost/benefit analysis to compare ONS current costs of statistical outputs (e.g. how much ONS spends on our Annual Population Survey and or other surveys), vs. potential costs involved in applying different approaches and or in modelling the financial approach to access commercial data.

3.3   Data standards and a 'national cell count' reference dataset

One of the primary challenges in working effectively with these data is that each operator collects, processes and applies methods differently to these data. Another challenge is to work towards a defensible sample size, and that ONS will require access to standardised data from multiple operators – so these methods and technological challenges are amplified. This approach will lay a foundation to overcome these challenges, define a standard for collecting, processing and potentially combining from multiple operators and will be a great

way to better build confidence in working with these data. A quick win goal will be to request operators collaborate to develop a 'national cell count' dataset. Such an approach could not only benefit ONS but could also support the UK and international statistical community to better utilise these data.

3.0 Timescales and resources

The below targets will likely need tweaking once the wider team has been consulted but the 6 month overall delivery timeline should remain a core goal. It also aligns to the research driven deadlines being adopted in the Data Science Campus so is a good bar to aim for. Appendix 2: Commercial engagement process provides a starting point to attach dates to.

3.1 Timescales

These are target outlines
- Week commencing (w/c) 18th April onwards: circulate 1.0 – 2.1 of this brief to Heads of Profession (DONE)
    - Secure and incorporate feedback and input to our approach (call scheduled with HSE, other input expected. (DONE – possible opportunity to enlist support and expertise from HSE)
- W/C 15th May: finalise ONS internal resources plan and budget for external resources
    - Agree and assign commercial services lead
    - Prepare public information notice / or single action tenders as is appropriate to develop data as a service and source necessary supporting expertise from the operator and consulting community (IN DISCUSSION)
    - Schedule supplier/partner engagement event
- **May 15th – 22nd May: shortlist and agree Challenge priorities with ONS leaders**
- **Secure communications input / external data privacy support**
- W/C 29th May secure external consulting support to run engagement
    - Proactively engage key partners to drive attendance to this event
- W/C 12th June share finalised three policy challenges and timescales to ONS and wider GSS community
    - Opportunity for any final queries or input
- W/C 26th June: Launch engagement process (Appendix 2)
- Mid-July: Conclude procurement of services
- Aug – Sep: work to prepare datasets and
- Oct: run Datathon
- Nov- Dec: Evaluate, report back learning, finalise data access requirements.

3.2 Resources

ONS will need to access resources across the organisation to participate in the following categories of work across this project and will need to leverage external expertise:
- DG and executive leadership (nominal time / chairing of Operator convened board)
- Commercial services (procurement and engagement)

- Data as a service (overall project management and coordination)
- Business development (business case and cost benefit analysis)
- Policy (finalising the data challenges)
- Communications (internally to Gov and GSS and externally to industry)
- Data science and engineering (DTM and Data Science Campus)
- Non-partisan mobile network data processing and methodology (external)

The level of commitment and resources will be more clearly communicated by MDR and
DaaS in a project plan when the project launches.

4.0 Engagement not just procurement

Building on lessons learnt from Flow of Funds it is essential ONS engage commercial data
partners at multiple levels and up to executive level (C-1 as a minimum level of engagement)
with organisations looking to provide commercial data as a service before, during, and after
the formal procurement of any service/s. This will ensure ONS is able to communicate the
benefits and opportunities the new legislation presents as well as the risks and pitfalls which
companies need to be aware of. It will also help ONS better manage public affairs
relationships externally – we already know these teams present a risk. It is inevitable we will
have to deal with data partners' sales teams to negotiate terms of service/s agreements but
access and support from their executive leaders should be built into any engagement model.
The Digital Economy Act and the corresponding powers it provides ONS, means our
approach to engaging these very commercial organisations needs to develop beyond the
traditional ONS purchase and supply style relationship with IT and data related vendors.

4.1 Commercial data as a service

This initial 6-month engagement is intended to provide a platform upon which mobile
operators and their key commercial partners can understand, support and help implement a
new sustainable model for commercial Data as a Service and that ONS is able to better
compare and benchmark industry expertise and capability. From UK led research,
internationally sourced experiences and recent ONS work carried out to date it is clear data
services in the following categories are of value to ONS:

- ➢ Processing data
- ➢ Devising and applying methods to analyse the data
- ➢ Communicating learning's and outcomes from the data
- ➢ Providing pre-processed data to appropriate service levels

Building on learning's from ONS Flow of Funds project, in no circumstances should this
initiative look to purchase data in itself – rather qualify and establish access inside
commercial boundaries that align to the Digital Economy Act.

5.0 Operational challenges

Some of the shortlist below was raised in our initial discussion at the Data Science Campus launch. We are sure there will be others:

- ❖ Human: joining up communications teams and messaging. Creating a clear consumer / citizen public good narrative around the use of mobile network data to support aggregate statistics and inform better decision making in the UK
- ❖ Technical: agreeing appropriate formats for bringing together data sets from different operators for different applications
    - ❖ E.g. you wouldn't typically attempt mobility analysis with counts (workplace population research done in CBS Netherlands disproves such an approach)
- ❖ Policy and practicalities: prioritising use case challenges
- ❖ External: gaining trust and support from those areas most likely to oppose or challenge our approach
- ❖ Prioritising use cases: researching, defining and prioritising the use cases. Qualifying and quantifying the socio-economic returns
- ❖ Ensuring open publication/dissemination of results
- ❖ Ethical approval – NSDEC?

Appendix 1: Understanding what the Act means for ONS commercial approach to accessing private sector data

When understanding the Digital Economy Act in relation to accessing commercial data (data from 'undertakings') wording is not the simplest to navigate. Explanatory notes in paragraph 201 of https://www.publications.parliament.uk/pa/bills/cbill/2016-2017/0045/en/17045en.pdf helps but he below is intended to help translate the powers of The Act into a narrative to help inform ONS commercial and legal approach to setting up data access with operators. (NB Ross Young to edit and approve before / if circulated externally):

Bill wording – the small print	Plain English version
(1) Subject to this section and section 45E, the Board may, by notice in writing to an undertaking, require the undertaking to disclose to the Board information which— (a) is held by the undertaking, and (b) is specified, or is of a kind specified, in the notice.	ONS can issue a notice to a company to mandate that they provide us information we outline in the notice – e.g. we could specify a data dictionary for processed data we know they hold or can obtain
(2) A notice under subsection (1) may require information to be disclosed on more than one date specified in the notice within a period specified in the notice.	We can state how often and on what dates the information should be provided across a timescale of our choosing
(3) A notice under subsection (1) other than one within subsection (2) must specify the date by which or the period within which the information must be disclosed.	Any notice we do issue MUST contain a date by which the information must be provided to us
(4) A notice under subsection (1) may specify	We can state exactly how we want data

Comment [w1]: MNOs might not agree to publication without their authorisation and this means only if positive results / not conflicting with commercial sensitivities etc.)

This was a big issue for Census who wanted ONS to have greater control over the ability to publish ….Official Statistics Code of Practice requires transparency etc.

Comment [w2]: NB different MNOs have different policies regarding opting out

the form or manner in which the information to which it relates must be disclosed.	to be provided – in what technical format, to what technical standard and how technically the data should be delivered (e.g. secure file transfer protocol SFTP)
(5) A notice under subsection (1) may require the undertaking to consult the Board before making changes to—(a) its processes for collecting, organising, storing or retrieving the information to which the notice relates, or (b) its processes for supplying such information to the Board.	This protects ONS against companies changing the way they provide us the data (common and devious practice in many proprietary software license models)
The following are not undertakings for the purposes of this section— (a) a public authority; (b) a micro business; (c) a small business. (14) Section 33 of the Small Business, Enterprise and Employment Act 2015 (definitions of small and micro business) and regulations made under that section apply in relation to subsection (13) as they apply in relation to subordinate legislation within subsection (1) of that section	Businesses of less than 50 people or €10m in turnover are **EXEMPT** from this legislation. They can legitimately charge us for data.
45E Further provisions about powers in sections 45B, 45C and 45D (1) Information disclosed under section 45B, 45C or 45D may only be used by the Board for the purposes of any one or more of its functions.	ONS can only use commercial data for statistics or research purposes not for anything else
Where personal information is disclosed under section 45B, 45C or 45D, the Board may not disclose the information to an approved researcher under section 39(4)(i) except with the consent of the person who disclosed the information.	If ONS is looking to use commercial data sources that contain data derived from personal data – e.g. Telco data sources or private healthcare data we MUST NOT disclose identifying attributes to any of our researchers. We must work with companies to anonymise rigorously from the start
45F Offences relating to notices under section 45C or 45D (1) A person is guilty of an offence if— (a) the person is given a notice under section 45C or 45D, (b) the person is required to comply with the notice, (c) the person fails to do so, and (d) the person does not have a reasonable excuse for that failure…. Digital Economy Bill Part 5 — Digital government Chapter 7 — Statistics 78 (2) A person is guilty of an offence if, in purporting to comply with a notice under section 45C or 45D, the person— (a) provides information which is false in a material respect, and (b) knows that the information is false in that respect or is reckless as to whether it is false in that respect.	Companies and directors and officers in companies we serve notices to can be legally charged if they do not comply with the notices! IN other words we will have teeth to pursue them with ☺

(3) A person who is guilty of an offence under subsection (1) or (2) is liable on summary conviction— (a) in England and Wales, to a fine, and (b) in Scotland or Northern Ireland, to a fine not exceeding level 5 on the standard scale. (4) If an offence under this section is committed by a body corporate with the consent or connivance of an officer of the body, the officer, as well as the body, is to be treated as having committed the offence. (5) In subsection (4) a reference to an officer of a body includes a reference to— (a) a director, manager or secretary, (b) a person purporting to act as a director, manager or secretary, and (c) if the affairs of the body are managed by its members, a member. (6) Where an offence under this section is committed by a partnership (whether or not a limited partnership) subsection (4) has effect, but as if a reference to an officer of the body were a reference to— (a) a partner, and (b) a person purporting to act as a partner.

Appendix 2: Engagement process and timescales

In an effort to communicate this as simply as possible have attempted a diagrammatical process flow – fully recognise this will need revising
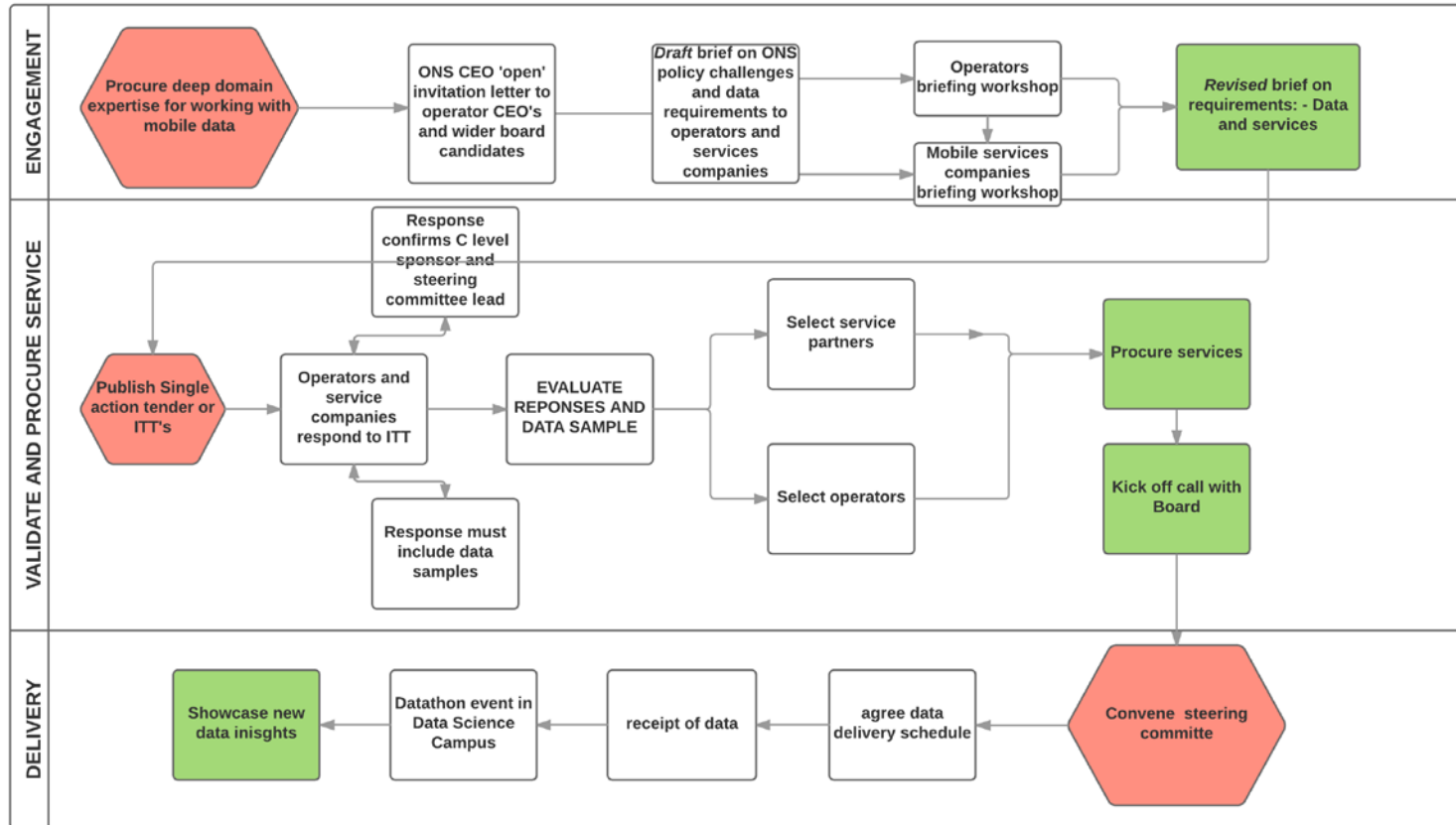
If it can mobilise swiftly enough ONS may want to consider implementing the DPS approach for commercial data acquisition proposed in *Faster, Better Decisions for Less.*

Operational structure

Steering committee should report to the board on a regular basis. Ideally the board would convene a short regular call with Board members able to deputise responsibility for that meeting to a direct report. This will make implementation easier and will support ONS to develop broad awareness for its work with the operators.

User committee

This team should be tasked with making it happen – working across the three core stages from engagement to delivery highlighted in Appendix 1. The team would likely be made up of internal and external expertise:
Operators:
  ❖ Technical implementation leadership from each operator
Technical and data standards/processing support:
  ➢ Independent technical governance, data licensing and or commercial expertise:
      ➢ Citi Logik (http://citilogik.com/ Steve Lees) and Flowminder
    http://www.flowminder.org/ (Andy Tatem)
        ➢ both bring relevant and different expertise
  ➢ Independent technical architectural support: Open Data Institute (Jeni Tennison)
Outreach (driving engagement from the external data scientist and research community)
  ❖ Datakind: http://www.datakind.org/our-story
Office for National Statistics
  ❖ Sponsor (Ian Cope)
      ➢ Technical and data architectural lead (TBC)
      ➢ Methods, Data & Research (Sarah Henry)
      ➢ Research and development (Tom Smith and team)
      ➢ Policy / Challenge use case sign off: (Emma Rourke and team)
      ➢ Commercial lead (Gary Evans)
Wider Government stakeholders:
  ➢ To be discussed but suggest x2/3 representative departments who have much to gain and input to the policy element.
      ➢ Possible representation from DfT, DCMS and or HSE may be useful?

Board

  ➢ This team should provide senior sponsorship and oversight to unblock any issues that may confront the steering committee in the project lifecycle. Chair: Office for National Statistics (John Pullinger or Heather Savory)
  ➢ Operator representation: C-suite leader from each operator and communications observer (e.g., public affairs representative)
  ➢ Privacy stakeholder lead: Sam Smith perhaps: https://www.linkedin.com/in/smithsam/
  ➢ Cabinet office or GDS?

Appendix 3: Statistical applications for mobile network data

Topic area	Sub topic	Statistical application
Population or call densities	Population estimates	Residential population
		Populations by time of day
		Seasonal population
		Ethnic population
	Urban planning or land use	Call density trends over space and time. "Hotspots" of activity
		Land use or activity (e.g. residential, rural, tourism, commercial, retail, leisure etc)
Mobility	Mobility behaviour	Prediction accuracy of mobility pattern prediction
		Social ties, link prediction
		Mobility patterns across different areas
		Mobility homophily (similarity)
	Urban planning	Home or work location
		Origin location of visitors to urban centre
		Exploratory analysis, call volumes, density areas or "hotspots" of activity
	Transport	Origin-destination flows for commuters, number of times per week commuting, commuting time and distance
		Rush hour transport pattern
		Main mode of transport in commute
		Ownership of car
Tourism		Feasibility study
		Tourist destinations
		International communities
Ethnicity and community		Ethnicity partitioning
		Regional boundaries
		Community detection
Socio-economic status and economic levels		Socio-economic status
		Economic levels
Infectious disease monitoring and prevention		Simulating disease spread
		Targeting humanitarian relief