

Figure I: Map of the pilot area



Stations included in the pilot:

Aldgate	Chalk Farm	Green Park	London Bridge	Piccadilly Circus	Tower Hill
Angel	Chancery Lane	Holborn	Mansion House	Regent's Park	Tufnell Park
Baker Street	Charing Cross	Kennington	Monument	Russell Square	Victoria
Bank	Covent Garden	Kentish Town	Moorgate	St. James's Park	Warren Street
Belsize Park	Dollis Hill	Kilburn	Mornington Crescent	St. Paul's	Waterloo
Blackfriars	Elephant & Castle	King's Cross St. Pancras	Neasden	St. John's Wood	West Hampstead
Borough	Embankment	Lambeth North	Old Street	Stockwell	Westminster
Camden Town	Euston	Leicester Square	Oval	Swiss Cottage	Willesden Green
Cannon Street	Finchley Road	Liverpool Street	Oxford Circus	Temple	

The thorny issue of tracking of location data without risking individual privacy is very neatly illustrated via a Freedom of Information (FOI) request asking London's transport regulator to release the "anonymized" data-set it generated from a four week trial last year when it tracked metro users in the UK capital via wi-fi nodes and the MAC address of their smartphones as they traveled around its network.

At the time TfL announced the [pilot](#) it said the data collected would be "automatically de-personalised". Its [press release](#) further added that it would not be able to identify any individuals.

It said it wanted to use the data to better understand crowding and "collective travel patterns" so that "we can improve services and information provision for customers".

(Though it's since [emerged](#) TfL may also be hoping to generate additional marketing revenue using the data — by, a spokesman specifies, improving its understanding of footfall around in-station marketing assets, such as digital posters and billboards, so not by selling data to third parties to target digital advertising at mobile devices.)

Press coverage of the TfL wi-fi tracking trial has typically described the collected data as [anonymized](#) and [aggregated](#).

Those Londoners not wanting to be tracked during the pilot, which took place between November 21 and December 19 last year, had to actively to switch off the wi-fi on their devices. Otherwise their journey data was automatically harvested when they used 54 of the 270 stations on the London Underground network — even if they weren't logged onto/using station wi-fi networks at the time.

However in an email seen by TechCrunch, TfL has now turned down an FOI request asking for it to release the “full dataset of anonymized data for the London Underground Wifi Tracking Trial” — arguing that it can't release the data as there is a risk of individuals being re-identified (and disclosing personal data would be a breach of UK data protection law).

“Although the MAC address data has been pseudonymised, personal data as defined under the [UK] Data Protection Act 1998 is data which relate to a living individual who can be identified from the data, or from those data and other information which is in the possession of, or is likely to come into the possession of, the data controller,” TfL writes in the FOI response in which it refuses to release the dataset.

“Given the possibility that the pseudonymised data could, if it was matched against other data sets, in certain circumstances enable the identification of an individual, it is personal data. The likelihood of this identification of an individual occurring would be increased by a disclosure of the data into the public domain, which would increase the range of other data sets against which it could be matched.”

So what value is there in data being “de-personalized” — and a reassuring narrative of ‘safety via anonymity’ being spun to smartphone users whose comings and goings are being passively tracked — if individual privacy could still be compromised?