Review of the TfL WiFi pilot

Our findings

TRANSPORT FOR LONDON
EVERY JOURNEY MATTERS

About Transport for London (TfL)

Part of the Greater London Authority family of organisations led by Mayor of London Sadiq Khan, we are the integrated transport authority responsible for delivering the Mayor's strategy and commitments on transport.

As a core element in the Mayor's overall plan for London, our purpose is to keep London moving, working and growing, and to make life in our city better. We reinvest all of our income to run and improve London's transport services and to make it safer, modern and affordable for everyone. We play a central role in delivering the Mayor's strategy to improve air quality and public health and to make transport accessible to all.

Our operational responsibilities include London Underground, London Buses, Docklands Light Railway (DLR), London Overground, TfL Rail, London Trams, London River Services, London Dial-a-Ride, Victoria Coach Station, Santander Cycles and the Emirates Air Line.

On the roads, we regulate taxis and the private hire trade, run the Congestion Charging scheme, manage the city's 580km red route network, operate all of the Capital's 6,300 traffic signals and work to ensure a safe environment for all road users.

We are delivering one of the world's largest programmes of transport capital investment, which is building the

Elizabeth line, modernising Tube services and stations, transforming the road network and making it safer, especially for more vulnerable road users, such as pedestrians and cyclists.

We work hard to make journeys easier through effective use of technology and data. We provide modern ways to pay through Oyster and contactless payment cards and provide information in a wide range of formats to help people move around London.

Real-time travel information is provided directly by us and through third party organisations, which use the data we make openly and freely available to power apps and other services.

We listen to, and act upon, feedback and complaints to constantly improve our services and work with boroughs, communities, representative groups, businesses and many other stakeholders to shape transport provision in London.

Improving and expanding transport in London is central to driving economic growth, jobs and housing throughout the United Kingdom. Where possible, we are using our land to provide thousands of new, affordable homes. Our own supply chain creates tens of thousands of jobs and apprenticeships across the country.

Contents

Foreword

Technology is transforming our lives, from how we work and enjoy our leisure time to the way we travel. For TfL, the dramatic pace of change has presented exciting opportunities.

Electric buses are improving air quality, while state-of-the art signalling means Victoria line trains run every 100 seconds. And smartcard ticketing is offering customers a wide range of convenient ways to pay.

The phenomenal rise in the number of people using smartphones and mobile devices, and the growing appetite for information on the go, led us to provide free WiFi at more than 97 per cent of our stations. We continue to invest in better digital customer information and making non-personal data freely and openly available to app developers.

And there are still untapped opportunities that, if explored, can radically improve customers' journeys.

Our transport system generates huge amounts of data. For instance, on a typical weekday, we collect 19 million smartcard ticketing transactions that offer a valuable insight into the number of people using our network. But this doesn't give us a complete picture – after they tap in, how do they travel?

Ahead of and during the pilot, posters kept customers informed

Which route or platform do they use? Do they take the first train or wait for a less crowded one? Do people choose the fastest route, or the most comfortable? How do customers move around our most complex stations?

A four-week pilot in November and December last year revealed that

depersonalised WiFi connection data can fill in the gaps, as it allows us to see how people travel beyond the gateline. It removes the need for costly, time-consuming surveys and means we can provide detailed customer information for specific times of the day, on individual lines, platforms and even trains. The potential benefits are enormous.

Given the success of the pilot, we plan to collect WiFi data connections across our network. We will keep customers informed, respect their privacy and clearly communicate how to opt out.

In this modern, data-driven world, the possibilities are endless.

Lauren Sager Weinstein
Chief Data Officer
Transport for London

Executive summary

Customers are at the heart of all that we do. Every journey matters to us and we work hard to make sure our transport network is safe, reliable and fit for the future.

More people than ever are using the Tube, with more trains running than at any time in its history. And we are continuing to transform the way we serve our customers. In addition to the substantial improvements under way on our trains and stations, we are looking to use technology to provide even better information to help people get around London.

Across the Tube network, 97 per cent of stations now have free WiFi installed, which helps people stay up to date while they travel. As a by-product of providing this service, our router network collects information about the connections made to the WiFi system. This data could potentially be very useful for understanding travel patterns.

From 21 November to 19 December 2016, we conducted a pilot to evaluate the usefulness of WiFi connection data. It involved collecting depersonalised data from mobile devices that made connections to our WiFi network at 54 stations. We then tested whether this could be used to provide benefits in four areas:

97%
of Tube
stations have
free WiFi
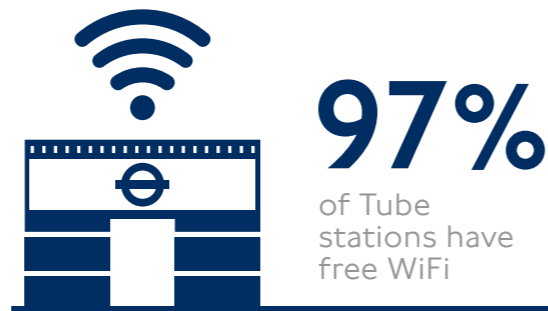
Customer information: could WiFi connection data help us provide better customer information for journey planning and avoiding congestion?

Operations and safety information: could understanding customer movements in stations help us deploy our people to best meet customer needs, and manage disruptions and events more effectively?

Transport planning: by better understanding how our customers use the Tube network, could we plan timetables, and our station designs and upgrades, more efficiently?

Prioritising investment: by measuring customer footfall and movements through and around stations, could we assess the effectiveness of our poster sites and retail units?

WiFi data can help us understand customers' travel patterns beyond the gateline

Historically, we have used a combination of data sources to understand customer movements between and within stations. This has included, since 2005, using depersonalised ticketing data to look at journey patterns. However, while this data is very useful for understanding gate-to-gate patterns, it does not tell us which routes customers take within stations.

Because of this, we have supplemented our ticketing data with manual customer surveys. This is time intensive, expensive and limited in detail and reliability. WiFi connection data, on the other hand, could provide a better and more cost-efficient solution. This is what we wanted to test in our pilot.

Over four weeks we collected more than 500 million depersonalised connection requests, and our technical process ran without error. Having collected the data, we needed to create new analytic tools and algorithms that could improve our understanding of travel patterns.

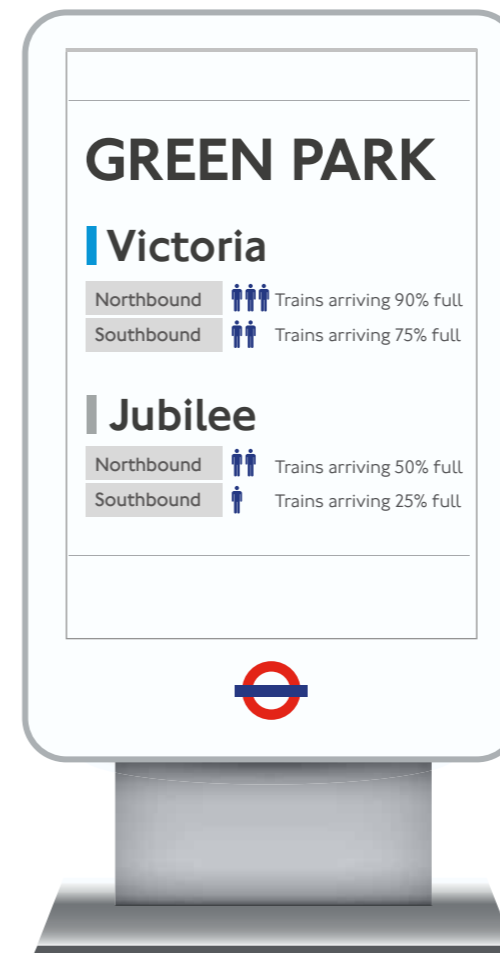From our data analysis, we have been able to conclude that:

• WiFi data can help us understand the paths customers take in stations, the platforms and lines they use, which route they take when many options exist and where they interchange

More than

500 million

depersonalised WiFi connection requests

• The aggregated data can show which sections of our network are crowded, at what times, and how this changes in response to events and network alterations

• We can use this data to power analytical tools and services that can improve the way we run and plan our network, and can provide our customers with much more detailed information

Further details and analysis on all these points are outlined throughout this report.

Openness and transparency is fundamental to the way we operate. From the first time we looked at whether we could use WiFi connection data to understand travel patterns on our network, we recognised the importance of making sure our customers clearly understood what we were proposing and why. We started with customer research and focus groups to test our proposal. Next, we assessed it through a structured Data Protection Impact Assessment.

GREEN PARK

Victoria

| Northbound | Trains arriving 90% full |
| Southbound | Trains arriving 75% full |

Jubilee

| Northbound | Trains arriving 50% full |
| Southbound | Trains arriving 25% full |

WiFi connection data could help us improve customer information for journey planning and avoiding congestion

We relied on guidance from the Information Commissioner's Office (ICO), the UK's independent authority set up to uphold information rights.

Ahead of the launch of the pilot we carried out a press and awareness campaign. We also pledged to share the results and to communicate our future plans for WiFi connection data.

The pilot was successfully delivered, using in-house expertise, for less than £100,000.

In view of the clear benefits to us and our customers, we are now planning to formally roll out network-wide WiFi data collection, so we can better understand travel patterns, provide enhanced information for our customers, and improve our planning and operations.

We will continue to inform our customers about any WiFi connection collection, the purpose behind it, and how we will use the data. And we will make sure people know how to opt out.

Purpose of the pilot

With 45 per cent of the Tube network in tunnels and stations as deep as 58.5 metres, there are many areas where cellular networks do not provide connectivity. In partnership with Virgin Media, we started installing WiFi in stations in 2012, and it is now available at 97 per cent of all Tube stations. This enables millions of customers to get online and find their way around using up-to-the-minute travel information.

Connectivity data is generated as a by-product of providing this service. We already use a range of data, such as aggregated and depersonalised Oyster and contactless payment transactions, to understand how customers travel across London. While this tells us where customers enter and exit the Tube network, it does not reveal anything about their route choices, such as which platforms and lines they used, where they interchanged, the paths taken to move through and around stations, and how they responded to network changes.
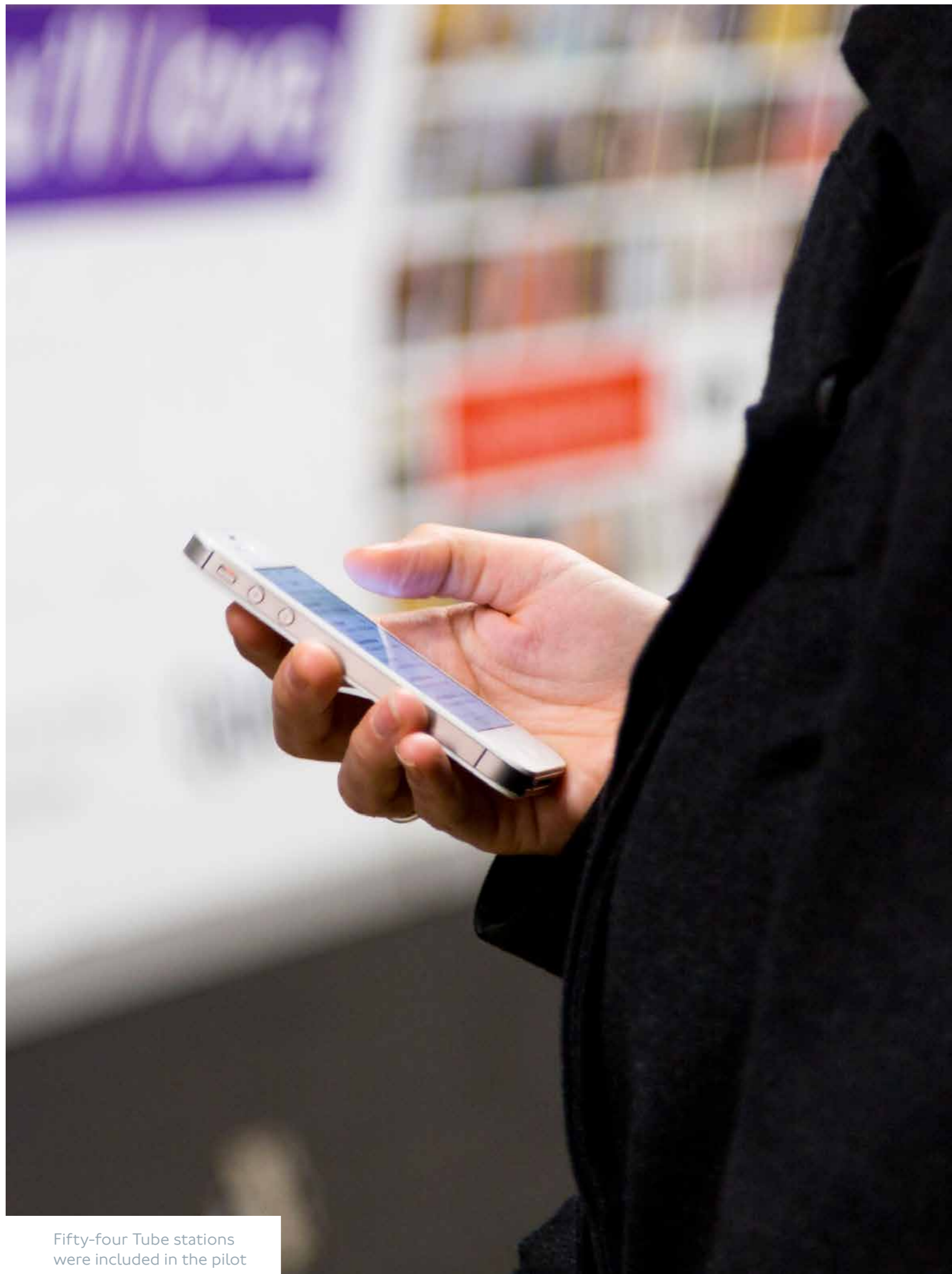
Owing to the layout of the Tube network, there are many route options that customers could take. To understand these choices and help us plan and operate the network, we have historically relied on paper surveys and manual counts, which are expensive, time consuming and limited in detail and reliability.

Figure 1: Map of the pilot area
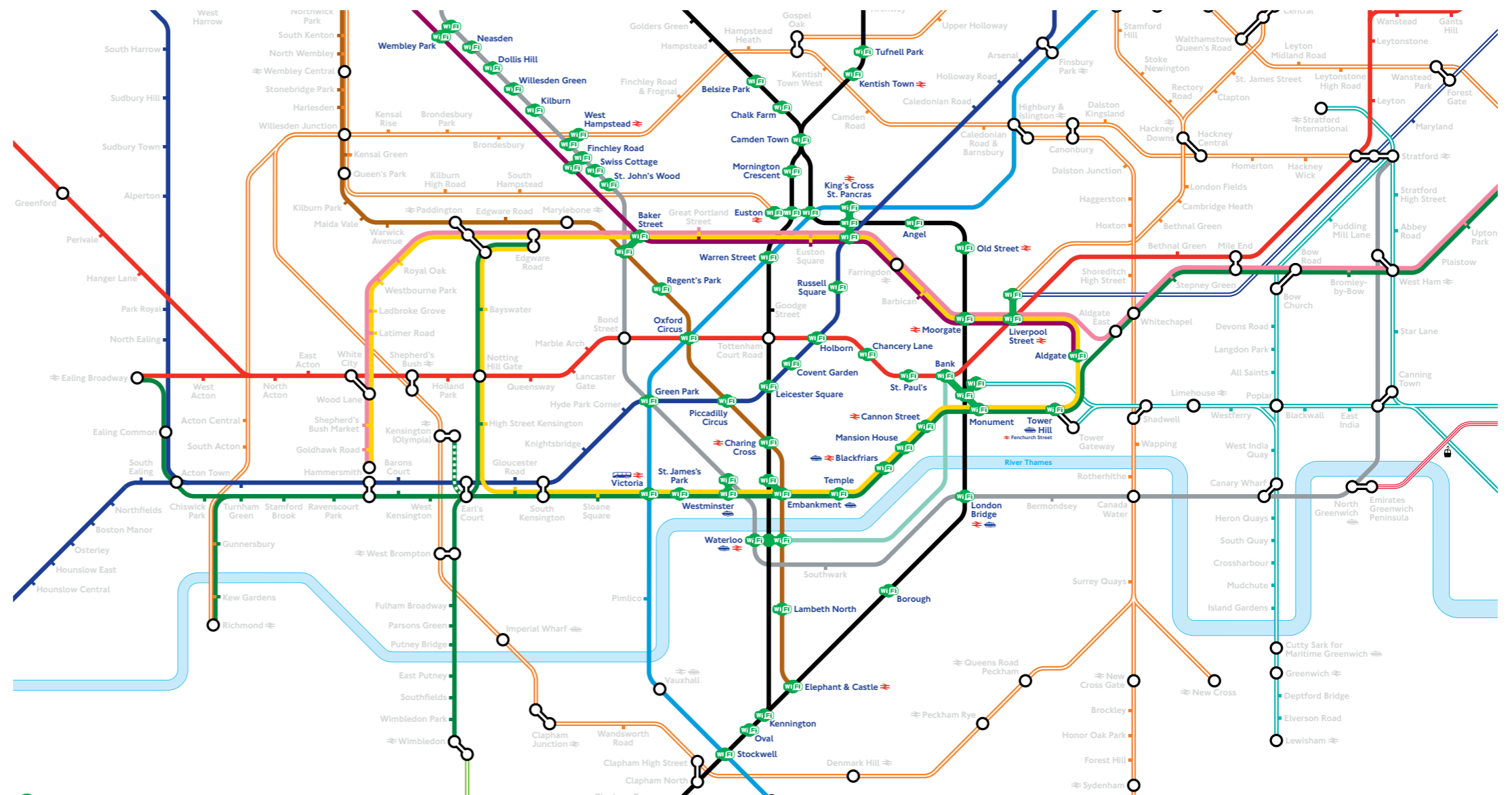
WiFi Stations included in the pilot:

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	Wembley Park
Borough	Embankment	Lambeth North	Old Street	Stockwell	West Hampstead
Camden Town	Euston	Leicester Square	Oval	Swiss Cottage	Westminster
Cannon Street	Finchley Road	Liverpool Street	Oxford Circus	Temple	Willesden Green

WiFi connection data could provide a greater understanding of crowding and collective travel patterns across the network, while also being more cost efficient.

We conducted this pilot to collect depersonalised WiFi data and analyse whether it could help us provide four important benefits:

Customer information: could WiFi connection data help us improve customer information for journey planning and avoiding congestion?

Operations and safety information: could understanding customer movements in stations help us deploy our people to best meet customer needs, and manage disruptions and events more effectively?

Transport planning: by better understanding how our customers use the Tube network, could we plan timetables, and our station designs and upgrades, more efficiently?

Prioritising investment: by measuring customer footfall and movements through and around stations, could we assess the effectiveness of our poster sites and retail units?

The pilot was carried out at 54 London Underground stations within Zones 1-4. Figure 1 on p10-11 shows the area covered.

Individuals who did not want us to collect the connections made by their device were advised to either turn off WiFi, or put their device into airplane mode while at one of the stations included in the pilot.

All TfL mobile devices issued to employees were excluded to avoid distorting the information being collected.

Fifty-four Tube stations were included in the pilot
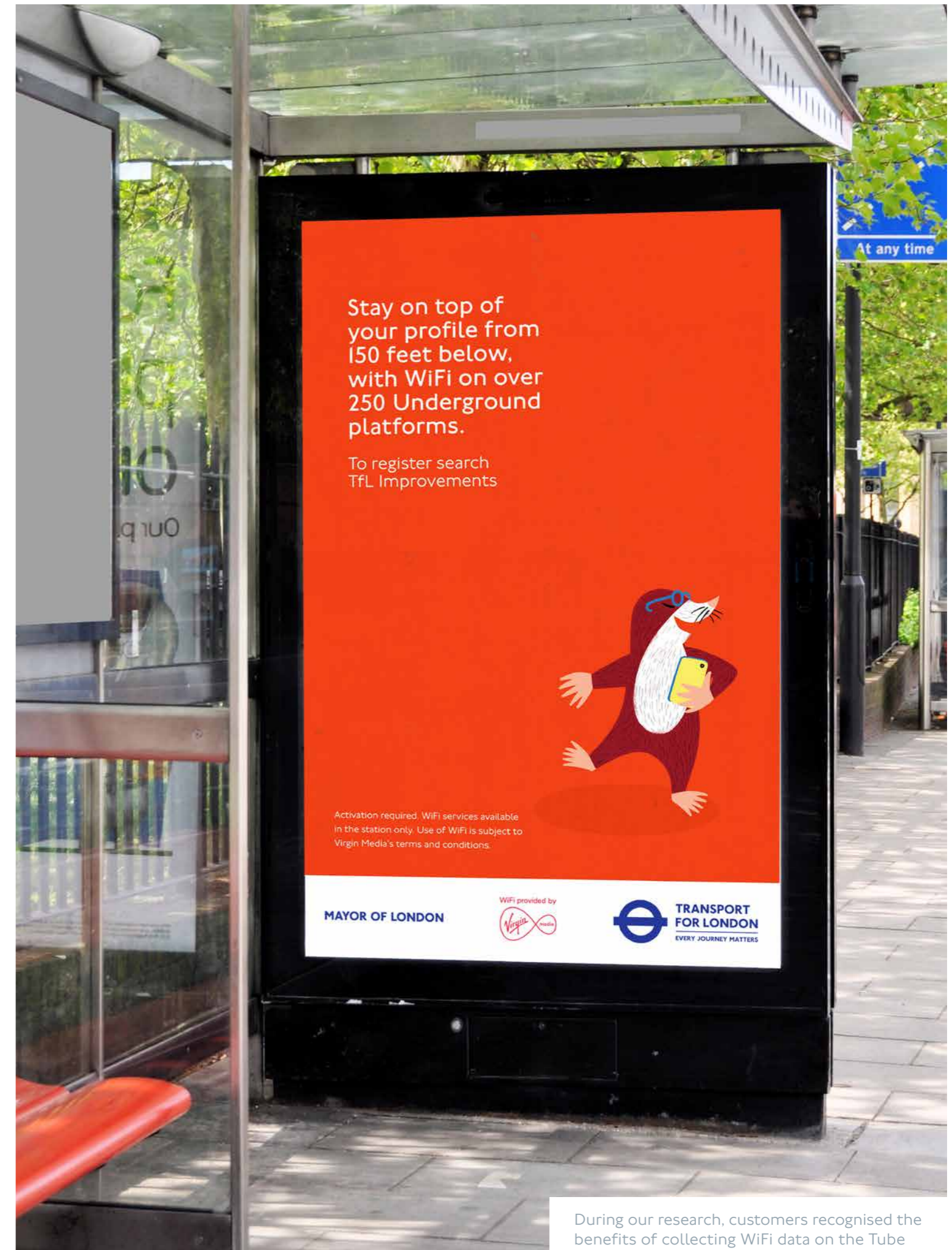
Customer research

From the point we first considered using WiFi connection data to identify travel patterns, we recognised the importance of making sure our customers clearly understood what we were proposing and why. To gauge people's views, we carried out customer research and set up focus groups. This tested a number of uses for mobile phone data to:

• Understand assumptions around the data we currently have and what we do with it

• Explore attitudes towards the different types of mobile connection data, both in general and distinguishing between types (Bluetooth, cellular, WiFi and GPS)

• Understand how customers feel about this data being used to improve journeys for them personally and across the system as a whole

• Guide how we could communicate our plans for using such data

• Help us understand the situations in which using such data was considered acceptable

We explored a number of options for using data generated by mobile devices. These are shown in Figure 2 on p16-17. The overall feedback was positive for all scenarios, particularly when our approach for collecting the data was transparent and offered clear and tangible customer benefits, with the ability to opt out.

Collecting WiFi data on the Tube was better received than other scenarios as it had a clear purpose and benefits. These findings helped us shape our pilot and highlighted that we needed to:
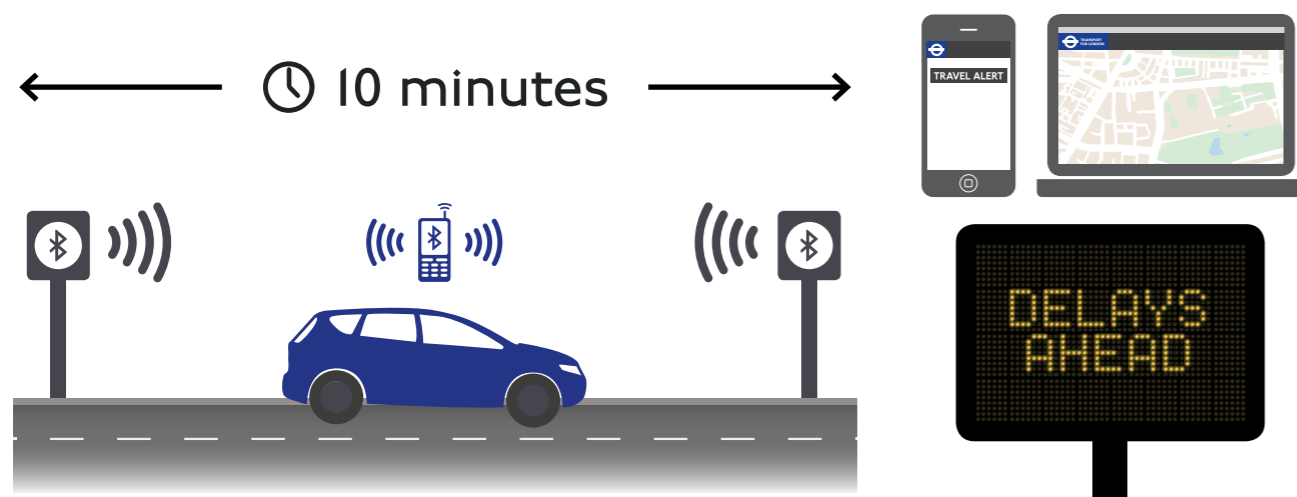
• Use a number of communication methods to inform customers of the pilot's purpose

• Communicate clearly that data would not be used to identify individuals
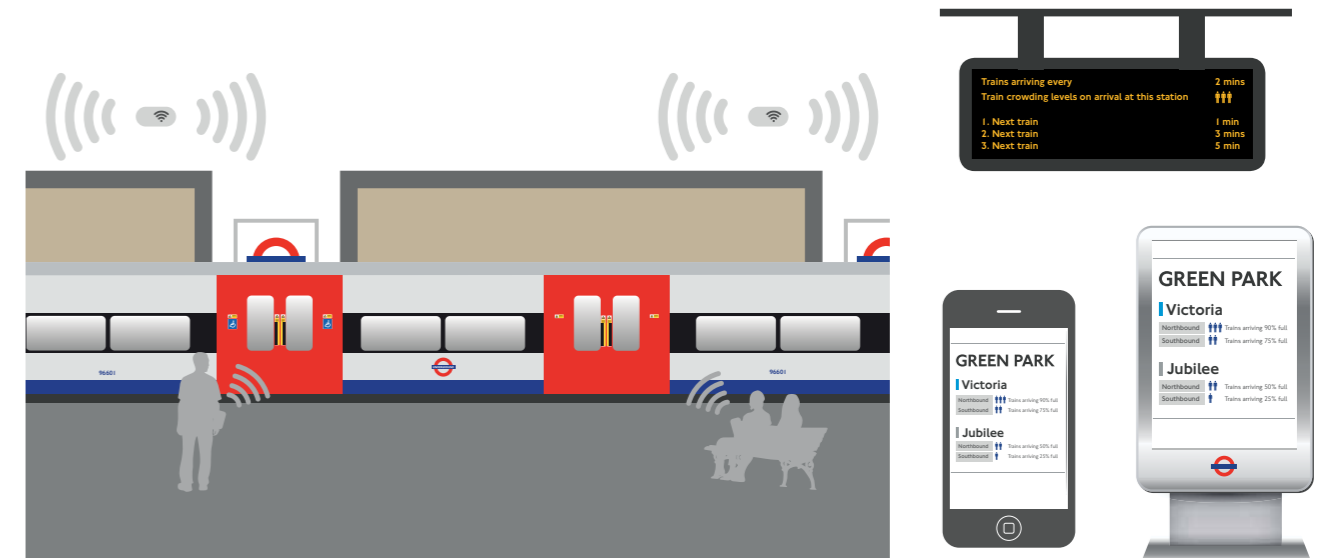
• Use the data to offer clear customer benefits

Stay on top of your profile from 150 feet below, with WiFi on over 250 Underground platforms.

To register search TfL Improvements

Activation required. WiFi services available in the station only. Use of WiFi is subject to Virgin Media's terms and conditions.

MAYOR OF LONDON   WiFi provided by Virgin Media   TRANSPORT FOR LONDON   EVERY JOURNEY MATTERS

During our research, customers recognised the benefits of collecting WiFi data on the Tube

Scenario I: Road beacons – for real-time traffic flow and congestion information

Scenario 2: Mobile phones – to understand movement and support planning decisions

Scenario 3: WiFi connection data on the Tube – for better customer information
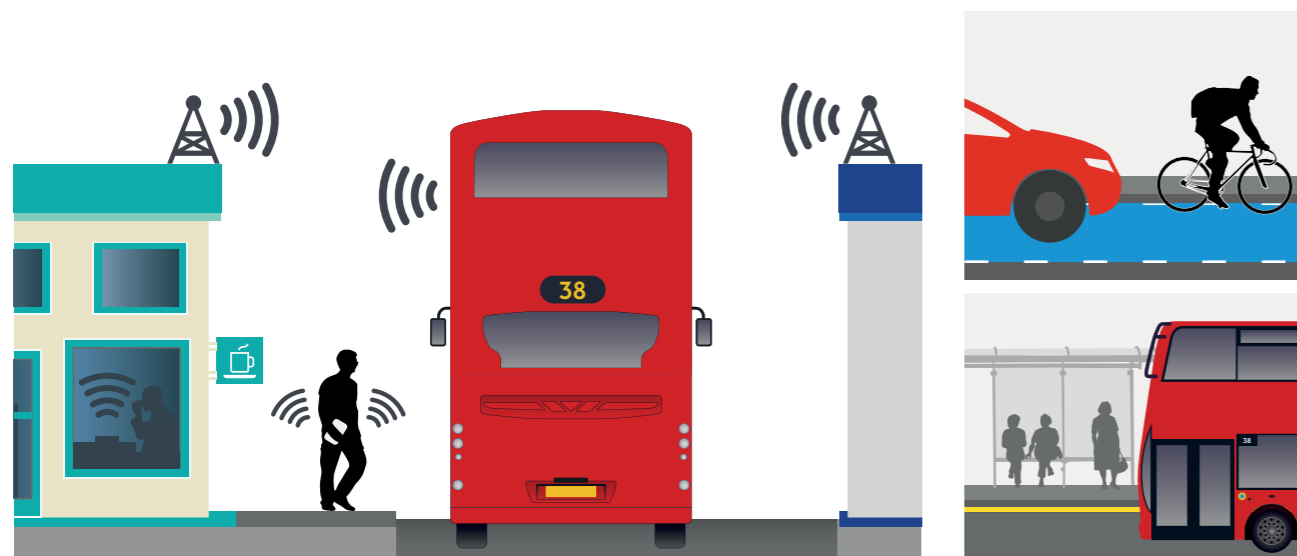
Scenario 4: Location data – for up-to-date information and personalised alerts

Keeping customers informed

We are committed to being open and transparent about how we use data. The ICO is the independent regulator of personal data in the UK. We considered the ICO's guidance on WiFi location analytics and its code of practice on privacy notices, transparency and control.

Consistent with best practice, we adopted the ICO's 'layered approach'. This involved using a number of communication methods to make sure customers knew that we would be collecting data and understood the benefits of doing so.

The week before launching the pilot, we issued a press release[1] that set out the benefits and scope of the project. This resulted in the pilot being widely reported in the media.

We publicised details on our daily travel page in Metro on 21 November (pictured right). The newspaper distributes more than 764,000 copies a day and is read by in excess of 1.9 million people.

Throughout the pilot, a new webpage was available on tfl.gov.uk/privacy with more information and answers to anticipated customer questions. Following feedback that this information was not prominent enough for mobile users who needed to scroll down, we made changes and displayed a link at the top of the privacy page for easier navigation.

More than 300 large posters were put up across the pilot area, including on platforms and at station entrances. They directed customers to our webpage for more information.

We used social media to inform people, including our 1.6 million Twitter followers and 386,000 followers of the TfL Facebook page.

Employees at stations and our Customer Contact Centre received briefing packs to help them answer customer questions. We published an article, similar to the press release, on our developer blog. Important stakeholders such as the Greater London Authority, ICO and trade unions were informed and offered more detailed briefings if required.

More than

300 posters

across the pilot area

---

1. tfl.gov.uk/info-for/media/press-releases/2016/november/wifi-trial-to-help-give-customers-better-journeys

What data was collected?

When a mobile device such as a smartphone or tablet has WiFi enabled, it will search for a WiFi network to connect to. This involves the device sending out a probing request that contains an identifier specific to that device, known as a Media Access Control (MAC) address. If a WiFi network is found that is known to the device, it will automatically connect. If the device finds unknown networks, it lists these in the device settings so the user can decide which, if any, to connect to.

During the pilot, if a device was near one of the 1,070 WiFi access points in the designated area, and it had WiFi enabled, we would have collected the request(s) to connect, even if the device did not subsequently do so. Once the device had connected, we also gathered these requests and authentications. No web browsing data or data from website cookies was gathered.

For customers who did not have WiFi enabled, no data was collected.

Figure 3: A fictional example of the data collected

Hashed MAC address	Date	Time	Access point
xIJx7F893LL4jO	01/12/2016	07:12:22	London Bridge 01
xIJx7F893LL4jO	01/12/2016	07:13:22	London Bridge 15
xIJx7F893LL4jO	01/12/2016	07:14:06	London Bridge 10
xIJx7F893LL4jO	01/12/2016	07:17:04	Bank 10
xIJx7F893LL4jO	01/12/2016	07:17:26	Bank 30
xIJx7F893LL4jO	01/12/2016	07:20:26	Old Street 06
xIJx7F893LL4jO	01/12/2016	07:20:50	Old Street 02
xIJx7F893LL4jO	01/12/2016	07:22:52	Angel 06
xIJx7F893LL4jO	01/12/2016	07:25:18	King's Cross St. Pancras 22
xIJx7F893LL4jO	01/12/2016	07:26:01	King's Cross St. Pancras 18
xIJx7F893LL4jO	01/12/2016	07:27:43	Euston 14
xIJx7F893LL4jO	01/12/2016	07:30:19	Camden 06
xIJx7F893LL4jO	01/12/2016	07:32:48	Chalk Farm 02
xIJx7F893LL4jO	01/12/2016	07:33:08	Chalk Farm 04
xIJx7F893LL4jO	01/12/2016	07:34:47	Belsize Park 04

More than

509 million

probing requests were collected at

54

stations from

5.6 million

devices
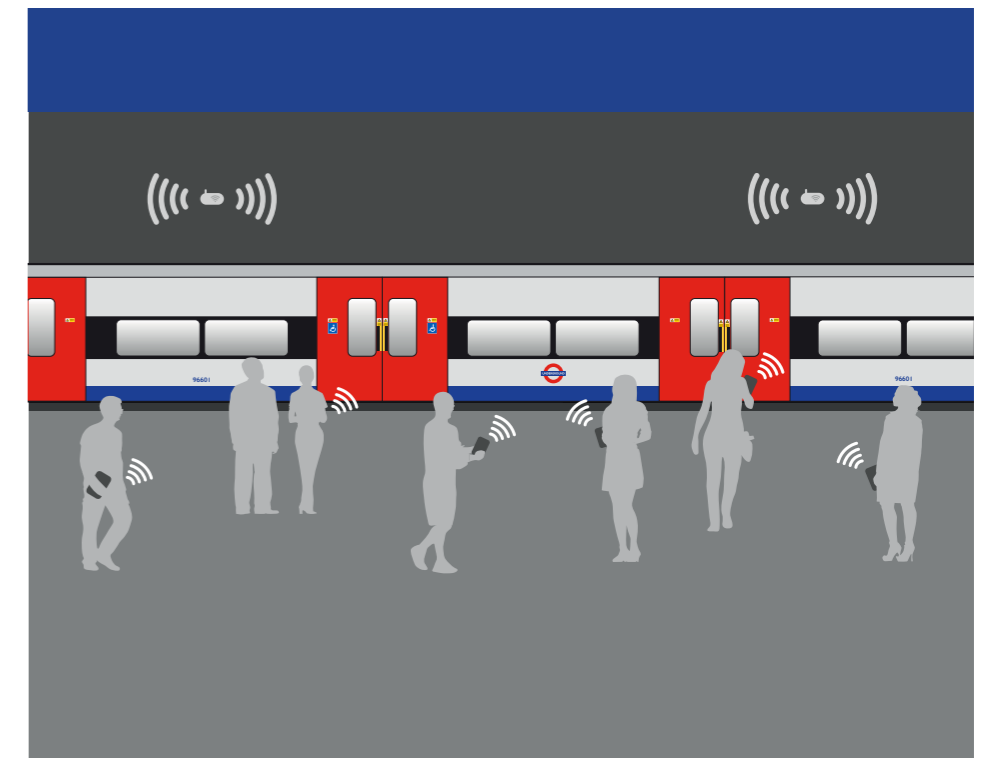
Throughout the pilot we only collected the minimum information required to test whether WiFi data could be used to provide the customer benefits we had identified. This included an encrypted, depersonalised version of the device MAC address, the date and time the device broadcast its MAC address, the access point it connected to, the device manufacturer and the device association type.

A fictional example of the type of data collected is outlined in Figure 3. It shows only the MAC address, date, time and access point for a device travelling from London Bridge, through stations on the Northern line (Bank branch) to Belsize Park. As we know where each access point is located (platform, ticket hall, etc), we can understand where in the station the device was when it connected.

Protecting customers' privacy and security

Protecting the privacy and security of our customers' data is of paramount importance and we recognise our responsibilities as a custodian of the personal data of millions of people. We understood that recording the location of a customer's device MAC address at a specific place and time could be considered as personal data. We employed a number of controls to make sure the pilot fully complied with the Data Protection Act 1998. We also briefed the ICO about our plans and met representatives to discuss these further.

Assessing our approach

To ensure our communications and data processing approach was appropriate, we completed an extensive Data Protection Impact Assessment before the pilot took place. It followed our formal TfL governance and project management processes.

Encrypting data collected

Each MAC address collected was automatically depersonalised (pseudonymised[2]) and encrypted to prevent the identification of the original MAC address and associated device.

In line with the ICO's WiFi Analytics Guidance[3], our irreversible encryption method included a 'salt'. This further enhanced security by adding an unknown character string to each MAC address before encryption. Even if a MAC address and the encryption mechanism is known, it is impossible for the original MAC address to be identified without knowing the salt. This reduces both the possibility of a single known MAC being identified in the data, and a look-up table of all possible MAC addresses being created and joined to the data.

The salt is not known by any individual and was destroyed on the day the data collection ended. Therefore, we consider the data to be anonymous and are unable to identify any specific device. As we cannot process known MAC addresses in the same manner as we did in the pilot, we are unable to complete any Subject Access Request[4] for the data we collected.

Data storage

The pilot data was stored in a restricted area of a secure server and has not been linked to any other customer information. Access was restricted to a small team that has regular privacy and data protection training. The data collected has not been, and never will be, shared with any third parties.

---

2. The process of distinguishing individuals in a dataset by using a unique identifier that does not reveal their 'real world' identity, in accordance with the ICO's Anonymisation Code of Practice

3. www.ico.org.uk/for-organisations/guidance-index/data-protection-and-privacy-and-electronic-communications

4. tfl.gov.uk/corporate/privacy-and-cookies/access-your-data
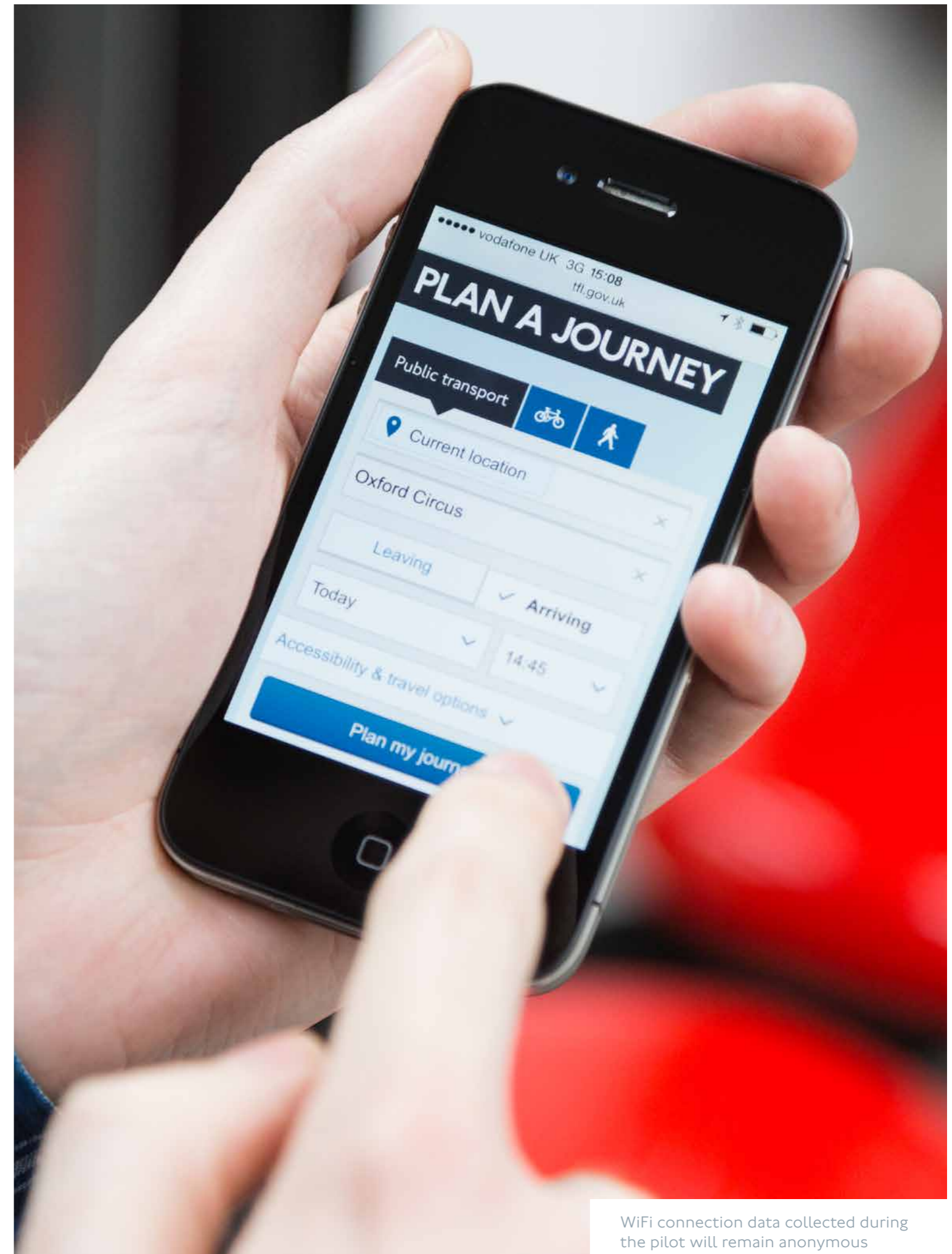
WiFi connection data collected during the pilot will remain anonymous

How successful was the data collection?

From the 54 stations included in the pilot, we collected 509 million probing requests from 5.6 million devices. King's Cross St. Pancras generated the most, with 37.6 million. The fewest were observed at Dollis Hill, where 10,000 were collected from seven access points.

Figure 4 shows the number of WiFi connections automatically collected on each day of the pilot. On average, more probing requests were generated on weekdays, when most people travel. Thursday 15 December saw the highest number of WiFi connections, with 21.2 million probing requests collected.

It typically took between 20 and 40 minutes from a device connecting to our network to collect the data in our secure server. Across the entire pilot period, the longest lag was 102 minutes between device connection and collection. We believe we can achieve real-time data collection based on what we learnt from the pilot.

Transforming WiFi data into journeys
To translate and interpret the data, we applied 'big data' techniques to convert these individual connections into a structured format.

Firstly, we linked individual connections from a device so that we could create 'journeys' – an end-to-end trip that is comparable to, but more detailed than, what we get from our ticketing data.
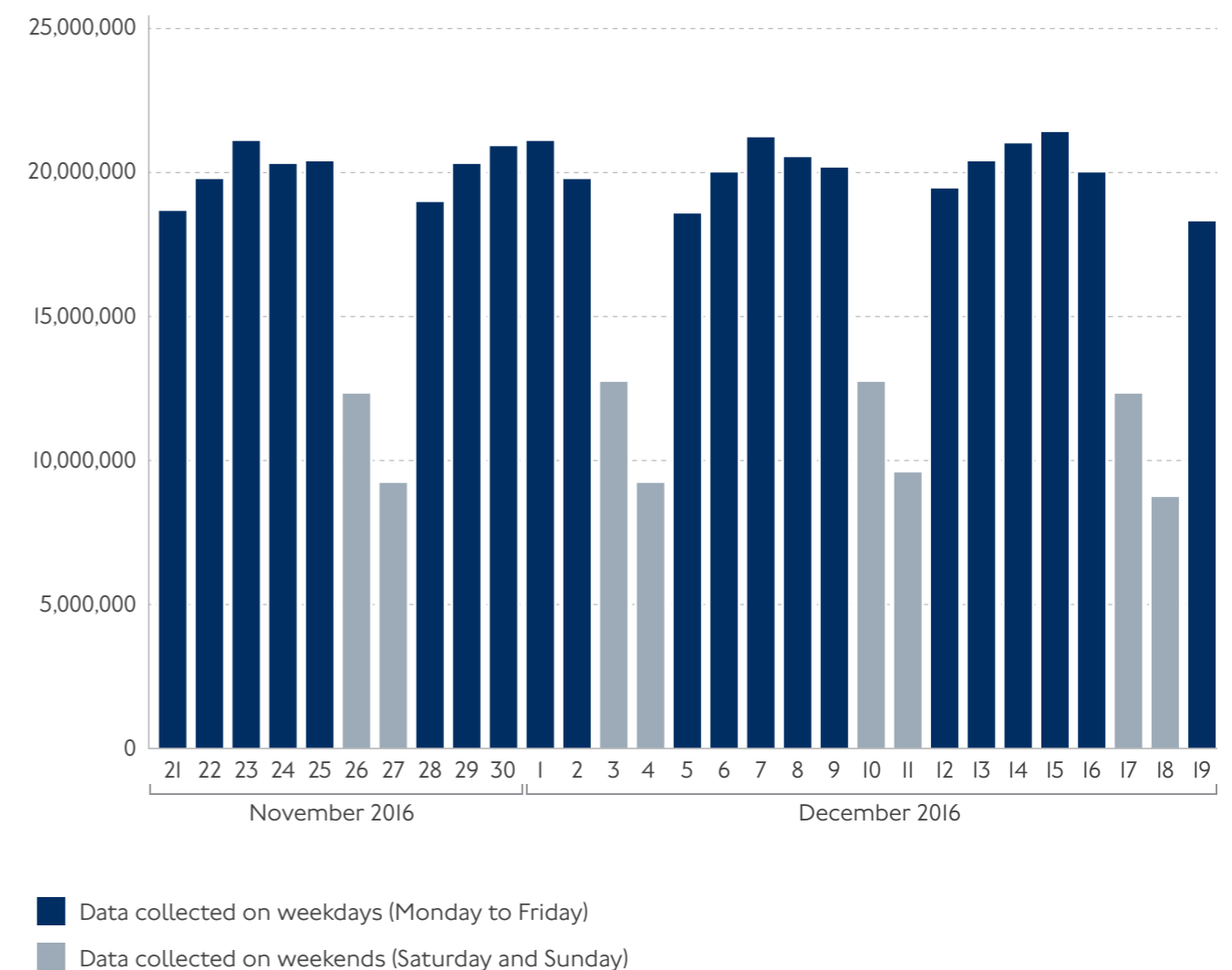
Using this approach, we constructed 42 million journeys from five million devices during the pilot. For some devices, we either collected single probing requests or the MAC address was randomised so we were unable to constuct journeys.

The next step was to break these journeys down into 'movement types'. Using advanced analytic techniques, we classified these as:

• Entry or exit – where a device was entering or exiting the Tube network at a specific point

• Pass-through – where a device was seen in a station as they passed through while on a train

• Interchange – where a device was changing from one line to another within a station

• Sub-categories of movements – where we determined that a device was boarding or alighting a train service

The reason for creating this classification system was not to track individuals or pinpoint the actions made by specific devices, but to understand, in total, the types of movements taken at different locations on the network.

Figure 4: WiFi probing requests collected each day

Legend:
■ Data collected on weekdays (Monday to Friday)
■ Data collected on weekends (Saturday and Sunday)

How data can help us improve our services

We carried out the pilot to test whether WiFi data could provide four potential benefits. Analysis has demonstrated that it would add considerable value and enable us to plan and operate our transport services more effectively. This includes providing better information and more detailed and timely updates on congestion so customers can choose travel options that best suit their circumstances. Our findings for each of the four potential benefits outlined on p13 are positive and these are explored in more detail below.

Customer information

WiFi data allows us to measure and understand crowding and demand on our network in greater detail than was previously possible. We can determine the paths customers take within and between stations. This, in turn, enables us to identify crowding levels in a station, on a specific platform, on a line or on a train at a given time of day.

Taken together, we can use this information to tell people about network conditions at every stage of their journey. This can help our customers to plan routes that best suit them, based on the network conditions at the time they are travelling.

The following insights from the pilot illustrate how this information may help customers improve the most relevant aspects of their journey:
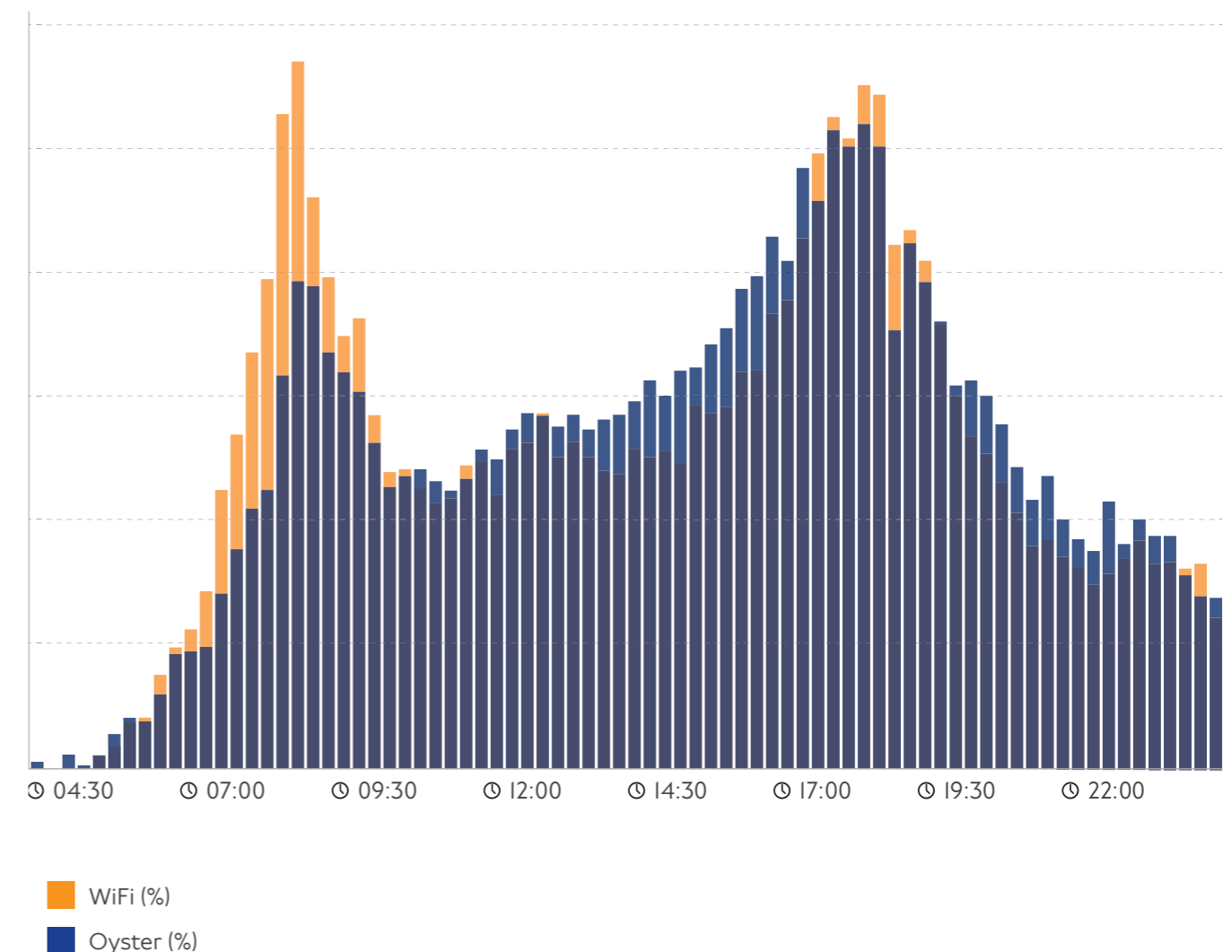
• When are individual stations, platforms and trains busiest?

• How busy are the different line and route options?
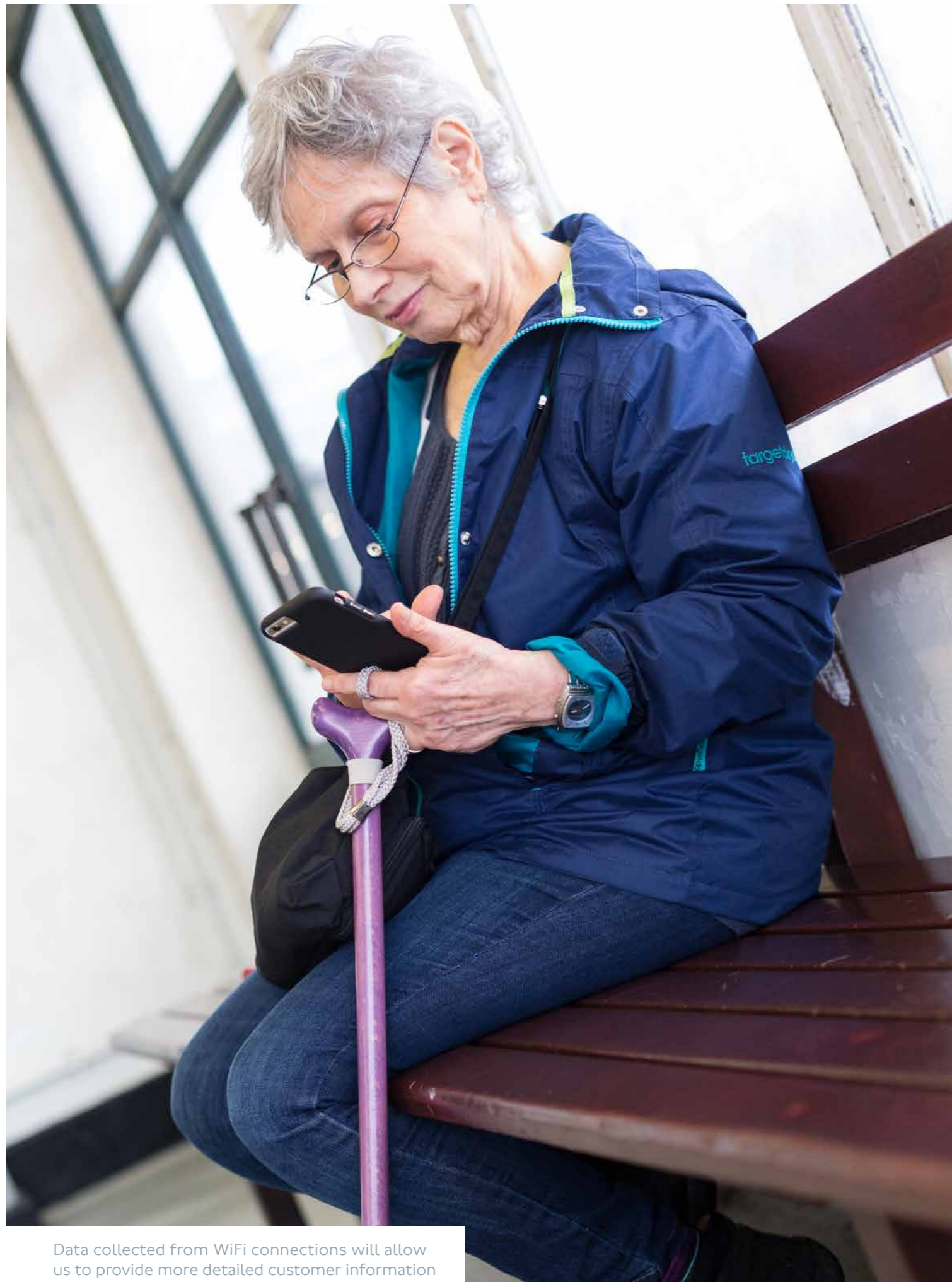
• How does disruption affect my journey?

Which times are busiest?

WiFi connection data is an accurate and reliable way to quantify crowding at a station at different times of the day. Unlike ticketing data, which only shows customers entering and exiting stations, WiFi captures interchanges. This is especially important when measuring crowding levels in large stations where many customers change between services.

Figure 5 highlights the demand at Oxford Circus station on a typical weekday during the pilot. The demand profile from Oyster ticketing data is shown in blue, and demand estimated from WiFi connections is orange. The differences can be attributed to Oxford Circus being a major interchange station where the Bakerloo, Central and Victoria lines meet. Ticketing data cannot show the sharp increase in crowding between 08:30 and 09:00 because it only captures customers entering and exiting stations.

Figure 5: Busiest times at Oxford Circus station, using ticketing and WiFi data

■ WiFi (%)
■ Oyster (%)

WiFi can also accurately capture crowding even when, on occasion, gates are left open, for example during large events or disruption. The information we currently give to customers is based on ticketing data and surveys for a typical day. WiFi would enable us to provide more accurate travel information on a continuous basis, under changing network conditions.

This data would be particularly useful to customers who would rather avoid crowding, even if it meant increasing their travel time. It would allow them to re-time and/or re-route their journeys to avoid the busiest sections of the network.

How busy are the different line and route options?
We can use WiFi data to understand where and when crowding occurs on trains and on specific sections of the Tube network. This information can tell our customers how busy different route options are for the same journeys.

To understand where crowding occurs on each line, we calculate the number of people on each train based on where and when devices connect to the WiFi network, in combination with train locations. Traditionally, this information has only been available for 15-minute intervals and is collected when we survey our customers on non-disrupted days.

However, this is relatively expensive and does not capture times when the network is disrupted. WiFi data will allow us to provide continuous, responsive estimates of demand on specific services, enabling passengers to make informed decisions about the journeys they make.

Data collected from WiFi connections will allow us to provide more detailed customer information

For example, someone planning to travel between Camden Town and St. James's Park is presented with two options when using TfL Journey Planner, as shown in Figure 6.

These options, however, do not help them make their decision based on factors other than estimated time and the number of interchanges.

Many customers want to understand how busy the train(s) and interchange station(s) could be so they know if they are able to board the first train. Data gathered from WiFi allows us to better advise customers by providing this level of detail.

Figure 7 shows how we could use WiFi data to improve the route information

available through our travel tools. It highlights the average demand across each section of these two journeys during the morning peak period (between 7:00 and 10:00). Assuming customers can board the very crowded train in the first option, this is quicker although significantly more crowded, especially on the branch between Camden Town and Euston, and on

the Victoria line between Euston and Green Park.

This crowding can be avoided by taking the Northern line (Charing Cross branch) direct to Embankment. Some customers may always prefer the quicker of the two options, but by providing these details we can help them make informed decisions.

Figure 6: Journey planning options for a trip between Camden Town and St. James's Park
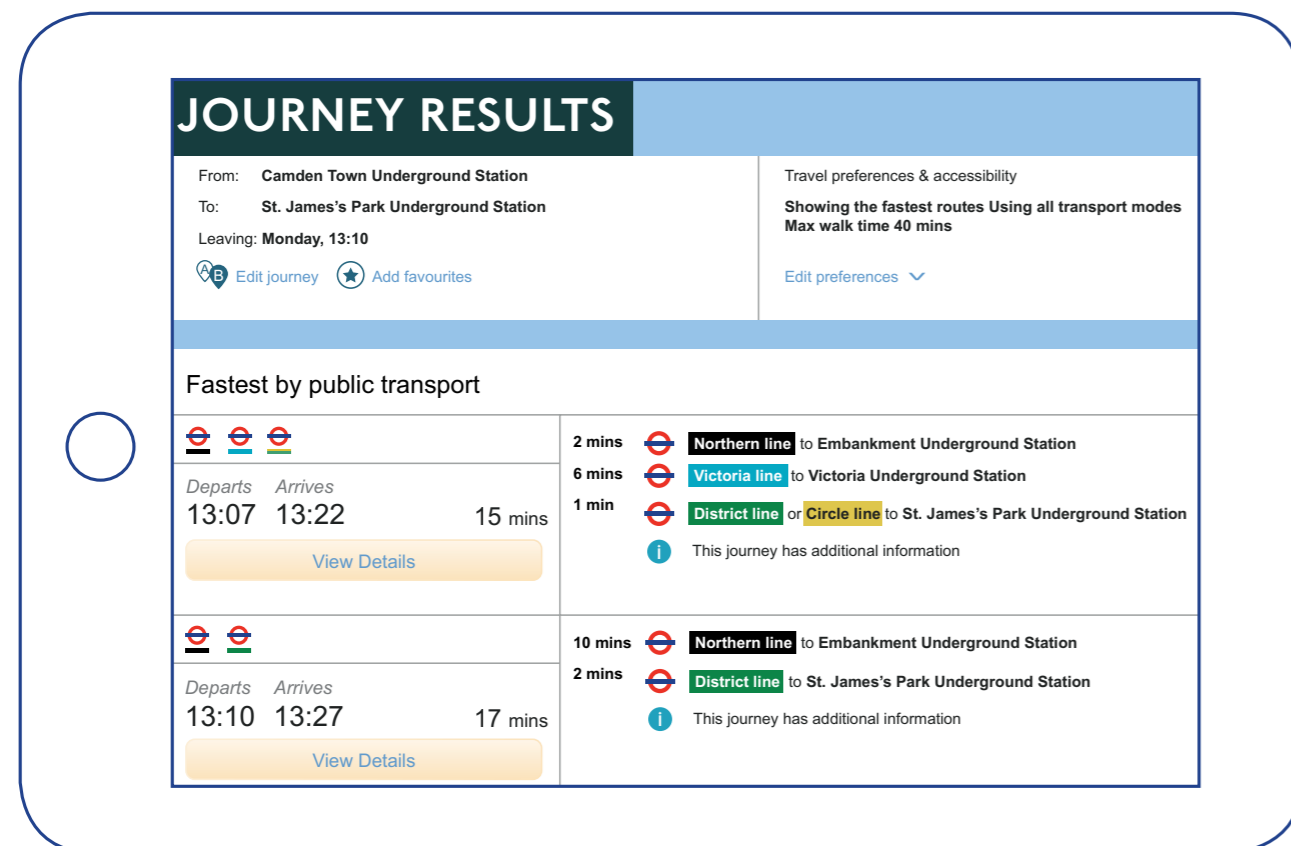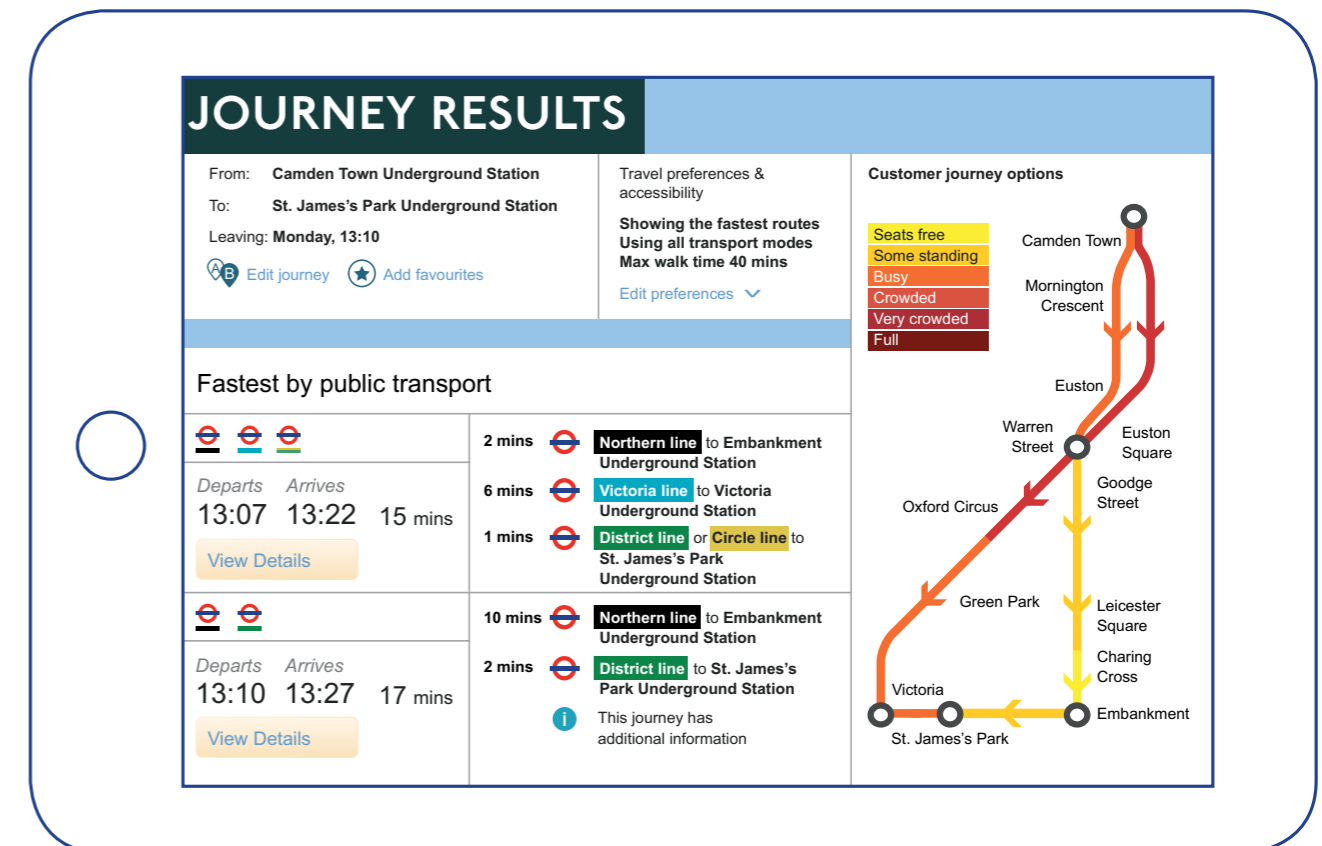

Figure 7: Enhanced travel tools using crowding information

In addition to being able to provide an insight into network-wide crowding levels, WiFi connection data could enable us to offer crowding information for individual trains.

While train-level demand remains relatively consistent across peak periods, there are some fluctuations in crowding levels between neighbouring trains. This is partly driven by the number of passengers on board when the train arrives, the amount of people attempting to board, and the number alighting from previous services who may still be working their way along the platform.

If we can provide information for crowding levels on specific trains, customers may choose to wait for the next one if it means a more comfortable journey. This would also improve the service for everyone by smoothing demand over the peak period.

Figure 8 shows the estimated number of passengers on southbound Victoria line services departing from Euston during the busiest times on the morning of 9 December 2016. The busiest services carry more than 1,100 people. Between 08:20 and 08:25, three trains departed with more than 1,000 customers, resulting in very crowded conditions. By contrast, trains departing in the preceding and following five minutes were around 10 per cent quieter, which represents around 15 fewer people per carriage.

We could use this information to tell customers how their journeys may be affected by crowding or congestion on the network. It could be presented at station and line level, allowing people to make decisions about the route they take. We currently provide information on the busiest times, but WiFi connection data would allow us to provide greater detail.

How long will my transfer take?
We can use WiFi data to understand the paths customers take within stations, and how long it takes to move around the station under different conditions. This gives us a detailed picture of the busiest areas at different times of the day.

Figure 9 on p34 shows how customers move from the Northern line (Bank branch) northbound platform to the Victoria line southbound platform at Euston station. Most people use the shortest path through the passageway. However, about a third head up to the main concourse then go down to the Victoria line platform. Five per cent take more complex paths. Those who use the passageway take around one to three minutes to travel between the platforms and those who use the concourse take around three to five minutes. This could be useful to less frequent travellers who want to understand transfer times and levels of crowding in stations.

Figure 8: Estimated number of passengers on southbound Victoria line trains departing from Euston, on 9 December 2016

Time	Passengers
7:50-7:55	100+ passengers
7:55-8:00	300+ passengers
8:00-8:05	500+ passengers
8:05-8:10	900+ passengers
8:10-8:15	700+ passengers
8:15-8:20	900+ passengers
8:20-8:25	Over 1,100 passengers
8:25-8:30	900+ passengers
8:30-8:35	700+ passengers
8:35-8:40	500+ passengers
8:40-8:45	700+ passengers

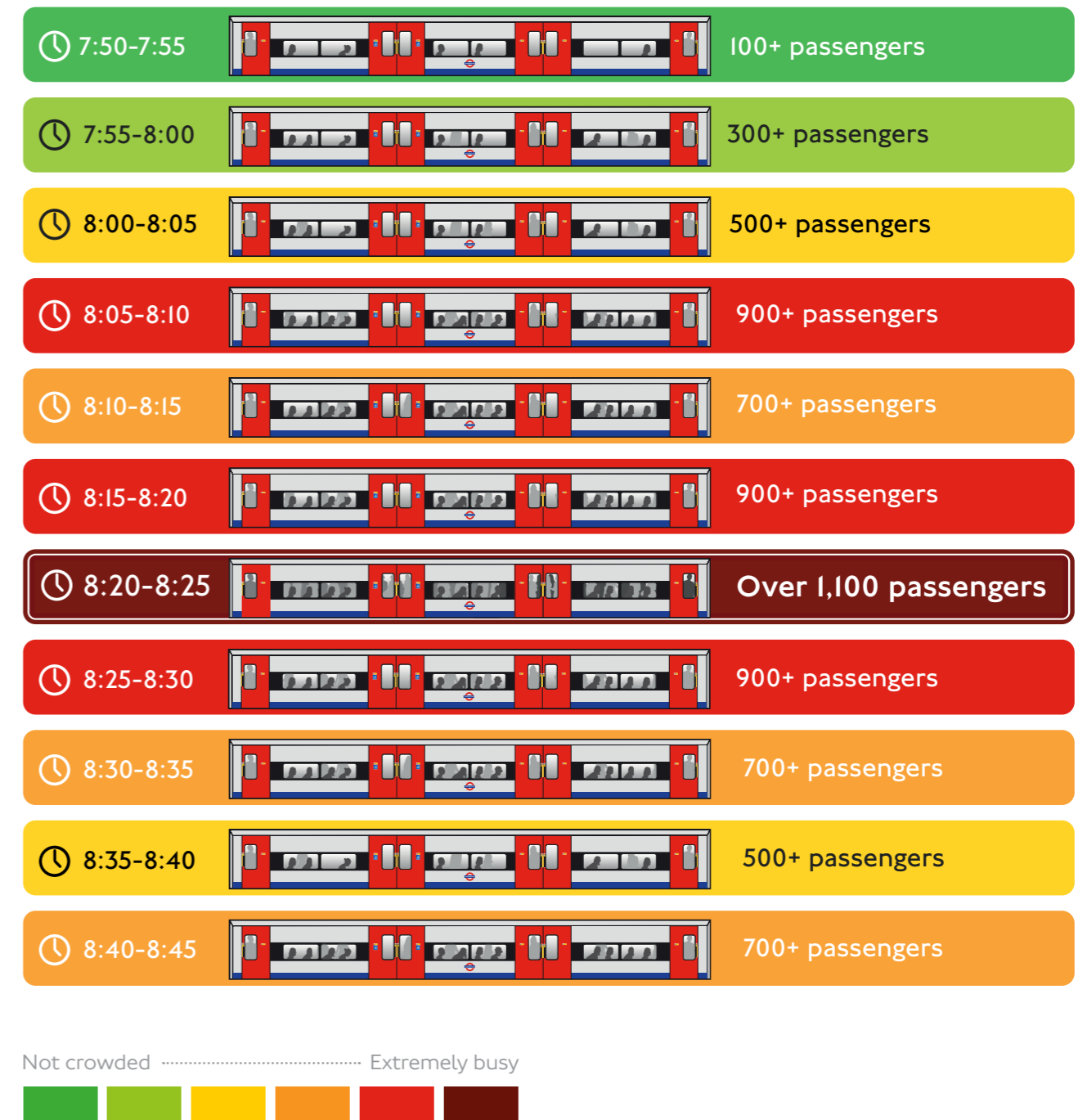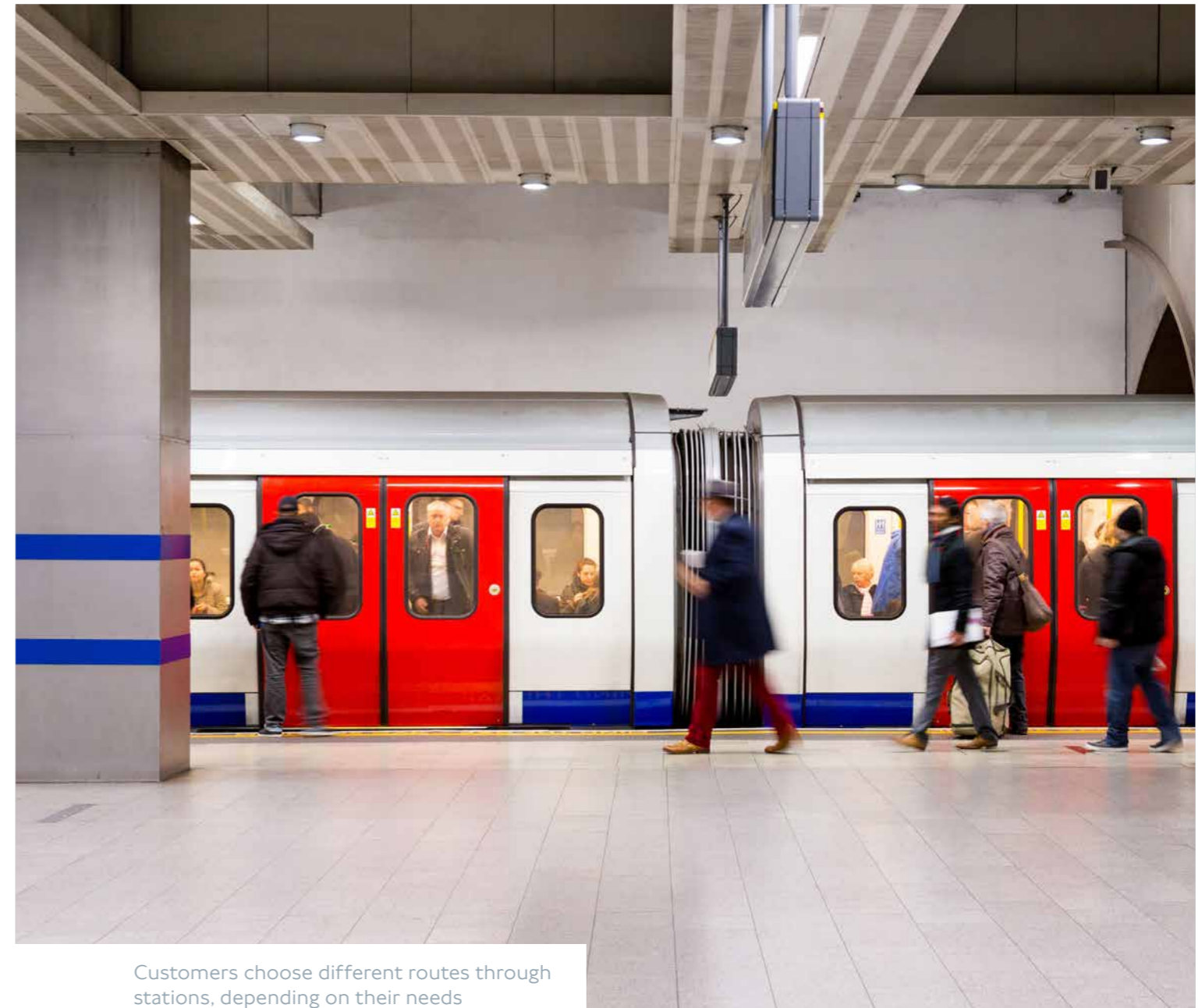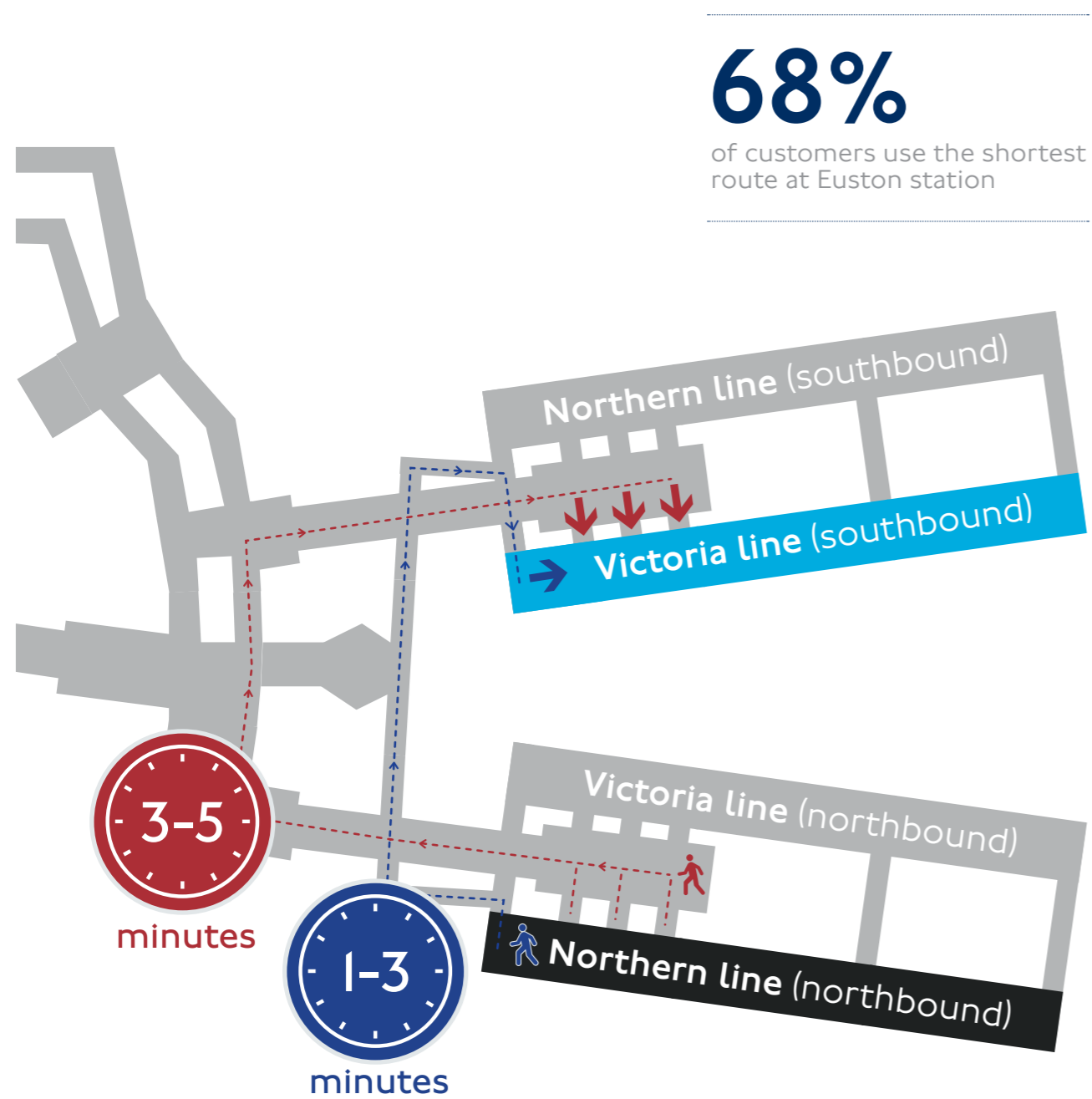Not crowded ···················· Extremely busy

Figure 9: How customers move from the Northern line (Bank branch) northbound platform to the Victoria line southbound platform at Euston station

68%

of customers use the shortest route at Euston station

Northern line (southbound)

Victoria line (southbound)

Victoria line (northbound)

3-5 minutes

1-3 minutes

Northern line (northbound)

Customers choose different routes through stations, depending on their needs

When choosing a route through a station, customers have different requirements and preferences. Detailed information about walk times and crowding can help people make the best decision for them. For example, someone with specific accessibility requirements, or passengers travelling with buggies or luggage, would benefit from knowing the least crowded path from one platform to another.

How does disruption affect my journey?

Another potential benefit is the responsiveness of the data to real-time events. Currently, a great deal of our understanding of crowding is based on surveys that only show a snapshot of the network when it is operating without disruption. We then make assumptions about likely customer responses, based on their typical behaviour.

Figure 10: Disruption that resulted from the Waterloo & City line suspension

08:00-09:00 Waterloo & City line suspended, which led to customers making a number of different travel choices to get to Bank station

Higher demand

Unchanged

- - - Suspended Waterloo & City line

Additional trains full of people using the route
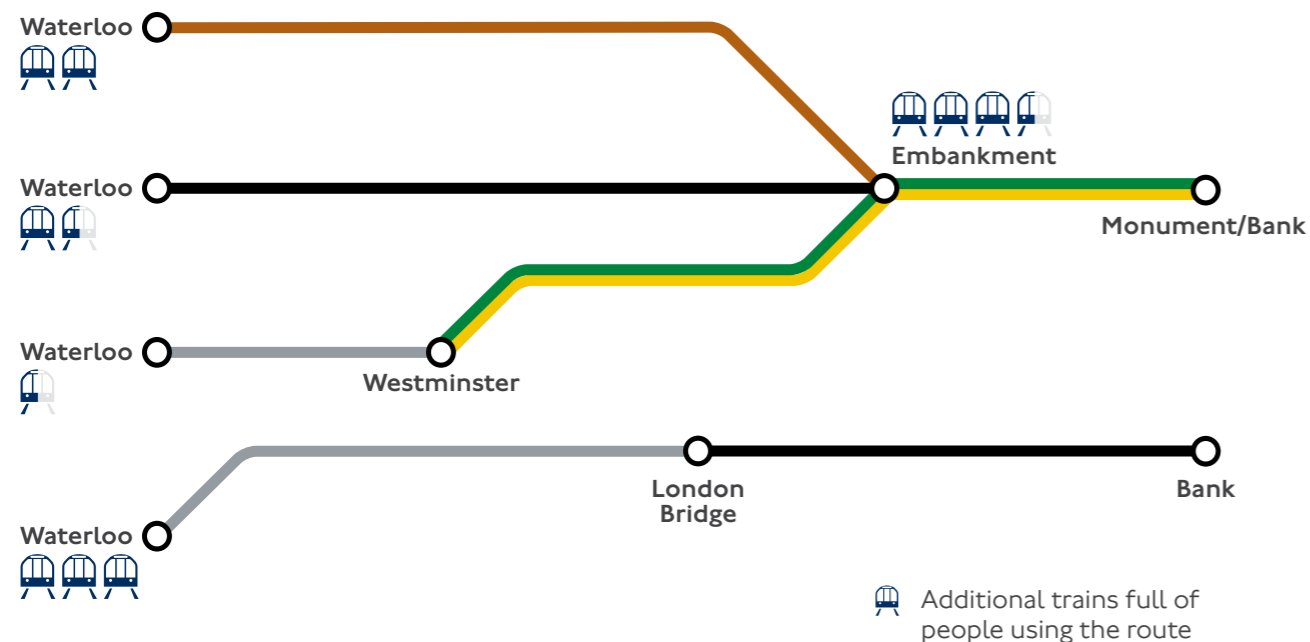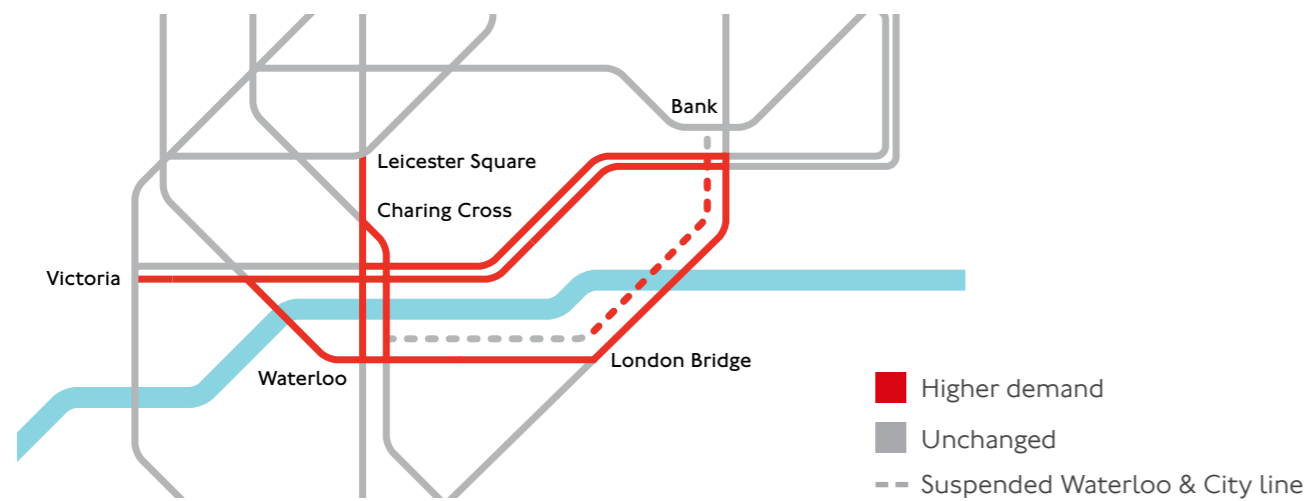
With WiFi data collection, which is continuous and offers a larger sample, we can better understand customer reactions.

On the morning of 9 December 2016, the Waterloo & City line was suspended between 08:00 and 09:00 owing to a signal failure. Under normal network conditions we would expect the line between Waterloo and Bank to be one of the most crowded sections on the entire network.

With the line unavailable, we saw customers make a number of different travel choices to complete their journey. Around 4,000 decided to use the Jubilee line to London Bridge, then take the Northern line to Bank.

In that hour alone, 3,000 more customers travelled from Waterloo on the Bakerloo line to Embankment, then continued on the Circle or District line to Monument. We also saw significant additional volumes of customers go north on the Northern line to Embankment and north on the Jubilee line to Westminster. All three of these responses meant a

transfer on to the Circle and District lines, which saw an additional 6,000 customers travelling east from Embankment towards Monument.

The impact of these various responses meant customers on the Northern, Bakerloo, and Jubilee lines from Waterloo experienced busier trains and interchanges than usual. Embankment station was particularly affected, with many more people changing here from the Northern and Bakerloo lines than normal – approximately 150 extra passengers for each train arriving – and many more attempting to board the Circle and District lines.

This data is valuable from both a customer and operational perspective as it enables a more informed response to changes in network conditions. It may help customers plan their journeys and allow us to look at the advice we give them, how and where we support them as they travel across the network, where we deploy our staff to offer assistance and how we operate our stations and trains to ensure the quickest and most comfortable alternative journeys.

Operations and safety

WiFi data could prove invaluable with regards to our operations and safety.

On 30 November 2016, Euston station was closed at around 18:40 owing to overcrowding. Figure 11 illustrates crowding in different parts of the station just before the closure. The darker areas are the busiest.

When comparing the amount of people in each area during the closure with typical conditions, it shows an unusual number of passengers alighting on the northbound Victoria line and Northern line (Bank branch) platforms, which

was likely to have been caused by an earlier closure at King's Cross St. Pancras station. As trains were not stopping there, customers chose to travel through Euston.

The result was many more people on the platform at Euston, causing severe congestion that meant we had to temporarily close the station to ensure our customers' safety. Figure 12 shows the dramatic increase in people on the platform on the day of the disruption, compared with the usual pattern.

The additional demand led to congestion in areas of the concourse and on the

escalators. This was reflected in the walk times between the northbound platforms and the ticket hall. Figure 13 shows the median time it took people to walk from the northbound Northern line (Bank branch) platform to the ticket hall. The time gradually increased from 18:10 as the station became more congested, up until it was closed.

Figure 11: Location of crowding levels in Euston before the station closure

Figure 12: The change in Euston platform activity prior to the closure

■ 30 November 2016   — Baseline

Figure 13: Increase in walk times between the platform and ticket hall

Walk times increased from **three minutes** at **18:00** to **more than 10 minutes** at **18:30** because of congestion
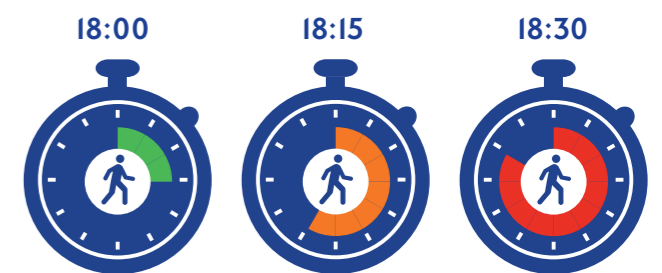
Analysing station congestion at this level of detail under real circumstances can help us investigate the causes of crowding, understand the effect conditions at one station have on other locations, and identify the knock-on impact of our operational decisions. It means we can improve the way we handle disruptions in the future and minimise the impact on our customers.

Transport planning

Much of the information collected and analysed during our pilot can also help our transport planners improve our network.

Our transport planning teams face a number of challenges, from upgrading the rail network to support a growing London, to looking at what our city will require in the longer term. To do this, they need to understand how customers currently travel and how they are likely to do so in the future.
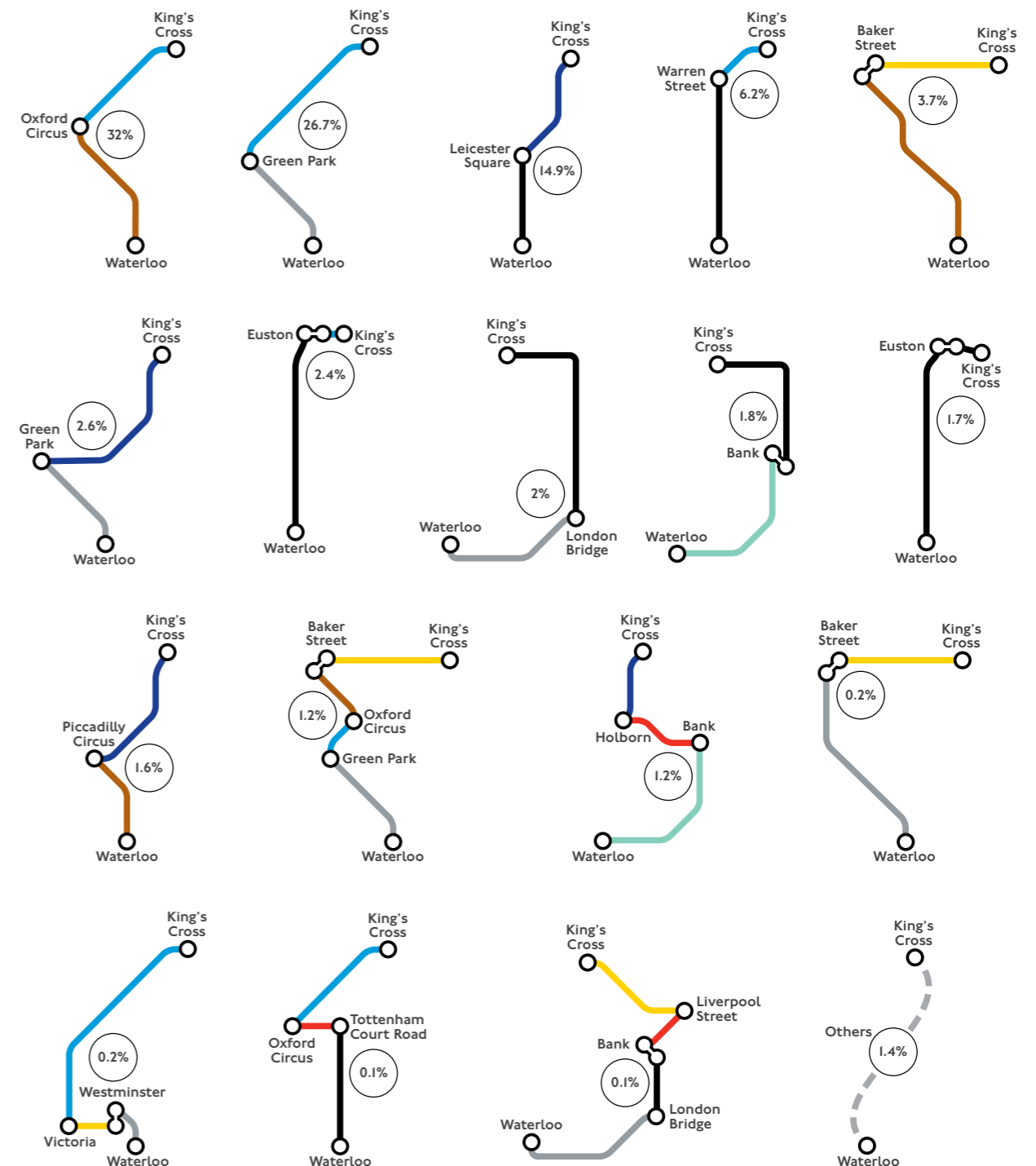
Using depersonalised ticketing data is a good way for us to understand demand, but it only provides an insight into gate-to-gate movements. For some areas, where stations are served by only one line, this is sufficient. However, much of our network is complex – with multiple interchange points and routing options, and many paths within stations. In these cases, our ticketing data cannot provide the detailed routing patterns required when planning our train services and station capacity.

To understand passengers' route choices, we have relied on customer surveys at a select number of stations each year. These ask people about their journey including when and where they started, which services they used, the lines they travelled on, and where they transferred between services. The results are then modelled to provide a detailed picture of demand. However, because surveys take time to run and process, we can only use certain stations.

This insight helps us to understand which stations and lines require capacity enhancements. We then run our transport models to predict how the network will operate with the proposed improvements in place. We can also predict the impact any works will have on other stations, which informs customer communications and operational plans. We use this modelling information to make changes to timetables so we can provide additional capacity and a more reliable service.

As part of our pilot, we tested whether WiFi connection data could provide our planners with timely information that would help them understand demand today, and shape planning and modelling in the future.

WiFi data can help us identify suitable locations for new retail opportunites, such as this coffee shop at Westminster station

Analysis has shown that the data provides a wealth of information on how customers use and interact with the network, both under normal conditions and during incidents and disruptions. For example, Figure 14 on p41 shows there are at least 18 route choices between King's Cross St. Pancras and Waterloo.

It also highlights the proportion of devices connecting to WiFi services on that route, using a single day's data. The complexity of paths chosen offers far more detail about customers' decisions than we get from our manual surveys.

The more detailed picture of demand that WiFi data offers has the potential to support most aspects of transport planning, from major station enhancements to smaller timetable adjustments.

Prioritising investment
WiFi data enables us to split stations into areas to understand the paths used to move between them, and the time it takes. We will be able to cut costs by either replacing paper-based surveys entirely, or by significantly reducing the number we need to carry out.

We can also use the data to get a more accurate understanding of where regular crowding within stations occurs. This can then help identify locations where further investigation needs to take place, particularly if crowding only happens at specific times or on certain days. By better targeting investment, we can provide customer improvements more efficiently.

In addition, WiFi data will enable us to provide aggregate footfall information to help generate increased advertising and commercial revenue.

For example, at Westminster station we recently opened a coffee shop on the eastbound District and Circle line platform. Before it opened, we carried out surveys and monitoring to understand the likely footfall so the retailer could decide whether the location would be a good investment.

WiFi data will allow more accurate, round-the-clock observations of potential sites by measuring the demand and dwell time at, and within, stations. This will allow us to better understand possible locations for future retail opportunities and help us get the best out of our assets so we can raise additional revenue to reinvest in the transport network.

Conclusion

By collecting and analysing WiFi data we will be able to understand aggregate customer movements at a given location and time. This can help companies that advertise at our stations to plan, monitor and measure their campaigns.

As the data will also enable us to identify the areas in our stations with the greatest footfall, we can understand which sites are most valuable to companies wanting to advertise. Using our digital advertising screens, where we have dynamic control over content, we can target advertising in the right place at the right time to maximise additional income. We will be able to frequently, and rapidly, measure how many customers could have viewed the adverts to enable companies to review their campaigns. And as WiFi can be used to split a station into many distinct areas, we can provide a detailed picture.

WiFi data could help advertisers target their

campaigns

more effectively, by showing which station areas have the greatest footfall

Our pilot has been a success both on a technical and analytical level. We collected 509 million anonymised probing requests in our secure data system, then created algorithms to understand travel patterns. The pilot showed that the connection data generated as a by-product of our WiFi services can offer substantial benefits for us and our customers.

Analytic tools and services built from this connection data can improve the way we run and plan our network, and can provide our customers with much more detailed information about their journeys.

In view of this, we are now planning to roll out a network-wide collection process. Given the importance that we place on transparency and openness, we will continue to inform our customers about WiFi connection data collection, the purpose behind it, and what it will be used for. We will also follow the ICO's guidelines and rules, and explain how customers can opt out by switching off WiFi on their devices.

We now look forward to taking the next steps to introduce this new technology to benefit our customers.