

Transport Committee

Oral evidence: [Future of transport data](#), HC 84

Tuesday 16 January 2024

Ordered by the House of Commons to be published on 16 January 2024.

[Watch the meeting](#)

Members present: Iain Stewart (Chair); Jack Brereton; Sara Britcliffe; Paul Howell; Karl McCartney; Grahame Morris; Gavin Newlands; Mick Whitley.

Questions 1–67

Witnesses

I: Dr Federico Botta, Senior Lecturer in Data Science, University of Exeter; Dr David McArthur, Associate Director for Urban Big Data, University of Glasgow; and Professor Susan Grant-Muller, Chair of Technologies and Informatics, Institute of Transport Studies, University of Leeds.

Written evidence from witnesses:

- [Urban Big Data Centre](#)

Examination of witnesses

Witnesses: Dr Botta, Dr McArthur and Professor Grant-Muller.

Q1 Chair: Welcome to today's session of the Transport Committee. This is the first oral session in our "Future of transport data" inquiry. To give the background, Select Committees spend a lot of valuable time scrutinising things that have happened, but we also want to look ahead at some of the upcoming opportunities and challenges. Last year, we ran a competition to select a couple of topics to look at. This is one of them. We are looking ahead at the future of data in the transport world. To begin, I ask our witnesses to state their name and institution and perhaps briefly summarise the areas of data policy they are currently looking at.

Dr Botta: Hello, everyone. My name is Federico Botta. I am from the University of Exeter, in the department of computer science, where I am a senior lecturer in data science. I am interested in issues around human mobility—how people move around and how we can use new forms of data, such as mobile phone data and other digital data—to understand that, and, related to that, how we can understand inequalities, both socially and geographically, around things like access to public transport, jobs, services and opportunities. More broadly, I am interested generally in how we can make data—transport data, in this context—more accessible and usable for policymaking and for society in general.

Q2 Chair: Thank you. Professor Grant-Muller.

Professor Grant-Muller: Thank you. I am Susan Grant-Muller, and hold the chair in technologies and informatics at the Institute for Transport Studies at the University of Leeds. I am a statistician by discipline and have been researching in transport for most of my career. I work a lot with pervasive and distributed technologies, sensor data, and understanding the movement of people, but also the uptake of new data and what that can tell us about people's choices. In particular now, that is not just data arising from the transport sector, but data arising from other sectors that is of use to transport.

Q3 Chair: Thank you. Joining us virtually from Glasgow, Dr McArthur.

Dr McArthur: I am David McArthur, reader in transport studies, based in the Urban Big Data Centre at the University of Glasgow. My interest is also in new forms of transport data, or data that can tell us something about transport and how that links with other urban systems—education, housing, the economy, health and so on.

Q4 Chair: Thank you all very much for your time today. As I said, this is the first session, so we want to explore the general landscape of data and what the potentials are. May I start by asking each of you a very high-level question? If we were to take full advantage of the types of data that are, or will become, available, how, potentially, could that change how we design, plan and deliver transport in this country? I will start with you, Dr Botta.



Dr Botta: There are many potential advantages that we could have in the future. We could, I guess, have better access to transport data and data that relates more broadly to transport. Something that is talked about a lot concerns better mobility services for people who have to travel around, people who want to commute to work or go to a place of leisure, or simply travel somewhere. Currently, a lot of transport is somewhat fragmented. You might want to take a train or a bus and then do some driving. Working it all out is not always trivial, depending on the particular area of the country, as there can be differences. Having a uniform or unified way of doing that would allow people to travel more easily, ideally spending less time travelling and potentially spending less money, if we could really understand the cost of journeys. That is currently quite difficult to do on an overall view. Obviously, all that has implications for reducing car use, which in turn has a potential impact on things like net zero. To me, the main opportunity is really around what is defined as mobility as a service, and a unified way of being able to travel, with systems that allow people to do that.

The last thing to say on this is that it is important, in particular, to think about it in terms of the users: not just the average person who might want to take a train or bus, or so on, but people with different needs such as disabilities or personal circumstances. They may not be able to travel in certain areas for reasons of safety or disability—all sorts of reasons. Really being able to understand and design services that can help with that is a huge challenge, and a huge opportunity where a lot of the data that, hopefully, will become available can really help.

Q5 **Chair:** Thank you. Professor Grant-Muller.

Professor Grant-Muller: Thank you. I won't repeat some of my colleague's comments. One of the main advantages is about putting users at the heart of the transport system, and improving efficiency and reliability and so on, but there is another angle that we need to think about. That is future resilience, and building resilience and preparedness for future shocks and crises of various kinds. Better data and sharing of data, and increased analysis of data, will help us to be better prepared and more resilient for what the future may hold for us.

Q6 **Chair:** Can I ask a supplementary? Is that, for example, using monitoring equipment to look at the resilience of bridges or embankments, or whatever it might be, so that, if a piece of infrastructure is predicted to weaken, an earlier or quicker intervention can be made, rather than waiting for it to fail?

Professor Grant-Muller: Yes, infrastructure is one interpretation of building resilience. Others would be health crises, which we have just had an example of, and deliberate interventions to the transport system. It is the sort of resilience that we need to think about in terms of peak demand for services where demand is significantly higher than the system is able to cope with. There are a lot of different aspects to



building resilience, and data can help us to look at the interactions between all of those and plan for a better system overall.

Q7 **Chair:** Thank you. Dr McArthur.

Dr McArthur: Yes, I agree with the points made by my colleagues. Some of the data availability goes to the traveller's experience, and how easy it is to understand how they can get from one place to another, perhaps in a healthier or more sustainable way. From an academic point of view, I tend to think about appraising and evaluating things. We often use data to try to understand what things work, what works well and how things might be changed. There are a lot of opportunities for us to access more data, and for it to be shared easily, so that we can understand better what sorts of interventions we might like.

On the resilience point, it is interesting that the pandemic highlighted some of the gaps. Many people suddenly realised that even quite basic data that we thought might be easily available, such as the number of people who might be on a bus or train, or walking down the street, is not easily available. It may not be collected, or it might not be shareable. We tend to get carried away with the more interesting forms of data, but even basic things like counts of who is where are sometimes missing.

Q8 **Chair:** Thank you. My colleagues and I will want to explore a lot of these issues in a moment, but I have one final question as an introduction, to try to quantify the opportunity. Given how we use data today, and how we potentially could use it, what is the gain to be realised in making transport safer, cleaner or more resilient, by whatever metric you want to use? What is the prize that we could realise?

Dr Botta: I guess there are a lot of advantages, and lots of things we could gain: one, going back to what I was saying, is around better, more inclusive transport for everyone, that really works for people across all demographic groups, across genders and different disabilities and needs. That is really important.

In the context of the UK, it is thinking about transport that works both for people who live in cities and for those in rural areas. It is a very different world, in a way. Transport is obviously a lot more prominent, frequent and available in large urban areas, compared with rural areas, so thinking about how data can help to improve transport across the whole UK and not just in large cities is one of the key points. It is also thinking about moving towards net zero, cleaner transport and less reliance on personal car journeys and car traffic, with increased reliance on public transport, which is used a lot in some areas and not so much in others. If we had better transport, putting users at the centre of the transport system, that would, hopefully, enable everyone to use it more, with big implications for net zero.

Q9 **Chair:** What I am trying to get at is whether we are looking at just small, incremental improvements in each area, or whether there is a massive



opportunity and a gap between how we use data now and how we could use it.

Dr Botta: For me personally, with the vision that I was just talking about, I guess the main gap relates to a unified way of looking at the transport system, where the data is all together. Currently, transport data and the related data that exists is very fragmented and in my opinion does not allow for an overall view. There is a real opportunity to have some kind of unified framework that allows us to study the whole system as one—not the bus and train sectors separately, and so on, but everything together in one place. That would allow us to take a big step forward on how public transport is designed.

Q10 **Chair:** Do either of our other witnesses want to comment?

Professor Grant-Muller: This is about a step change. Technology changes and improvements, and increasing pervasiveness of sensors for all kinds of reasons, give us an opportunity to make the joined-up picture across the whole system that for various reasons we have not quite had before. We need that boldness to be able to step forward and look at data in quite an ambitious way, to gain an overall integrated picture of what is happening and how we can better prepare the transport system for the future.

Q11 **Chair:** Dr McArthur.

Dr McArthur: Yes, I agree with those points. The interesting thing about transport is that it is connected to absolutely everything. You cannot understand housing or the economy without thinking about how transport plays into it. The ability to take a holistic view and understand how to get the most bang for your buck, as it were, can be transformational.

Chair: Thank you.

Q12 **Grahame Morris:** It was interesting, Dr Botta, that you mentioned MaaS—mobility as a service. The Select Committee did quite a detailed investigation and published a report, and I thought the potential of MaaS as a concept was fascinating.

To go back to the terms of reference of this inquiry, what are the biggest gaps in available data about transport networks and travel at the moment? I don't have a car. I rely on public transport—buses and trains—and one of my bugbears is real-time bus information. It is fantastic in London. It says, "The No. 2 bus will be here in two minutes," or, "The 136 will be here in five minutes." Where I live, we have, in theory, a similar system, which doesn't work. If this is about empowering individuals to make decisions—it might be about encouraging modal shift, and that might be a good thing—surely we need accurate real-time information. What are we not collecting, and where are the gaps currently in the system?

Professor Grant-Muller: I understand the sentiment you have just expressed. I have a similar experience in my location. It is not enough for



us just to collect or share lots more data. It is about making the data services—the data ecosystem—work better, more reliably and more accurately. That needs all sorts of things. It needs investment in data and in ICT systems; the hardware, and so on. It needs the back-up—the people who can troubleshoot when things go wrong—and the connectivity that shows that something has gone wrong with the bus stop information system so something can be done about it. The investment and probably the agenda are not just about data per se, but a data ecosystem and everything that goes with it.

- Q13 Grahame Morris:** Are there, Professor Grant-Muller, any specifics? I gave the example of the real-time data at bus stops, but are there any particular gaps that you can identify in relation to motorists or the rail system, and say, “If we did this, this would be possible”?

Professor Grant-Muller: For me, there is one very clear gap, which is around the demographic. Who, exactly, is travelling where? From where to where, for what reason, at what time of day, and what are their characteristics? What is their age, class, gender, and income group? That is a huge gap in what is out there at the moment. If we had that information, we would be able to create far more tailored services. We would understand the needs of, say, older people or those with particular kinds of needs around travel. We would be able to identify where there is a need for levelling up. We talk about levelling up in terms of geographies but, actually, it is about sub-parts of the population as well. I could go on and on, but I won't. That, for me, is the missing data at the moment. Operators collect it periodically, and they do ad hoc surveys, but there are sensitivities about how the data is collected—where it comes from, how it is protected, how it is used. That is the big piece of the jigsaw for me.

- Q14 Grahame Morris:** Dr McArthur, may I ask whether you have any particular views or examples that you could share with the Committee?

Dr McArthur: It is an interesting point. Some of the data may exist or be collected, but the ownership is often quite fragmented. In Glasgow we have several bus companies. Some of the buses have the equipment for real-time tracking, and some do not. Some will provide it, and some may not. There are commercial sensitivities about whether they want competitors knowing which routes are successful or what times of day they should be running services. That can pose barriers to transport planners.

Then you have other actors. Perhaps Google may have a better idea of how many people are on buses than even some of the bus operators. At Google they gather large amounts of data that they then control. You end up with little pots of data sitting in different places all trying to cover the same thing. The local authority may have done a travel survey. The operator of the regional transport partnership may have survey data connected to it. That makes it quite difficult, although from each actor's perspective you can understand why they do not necessarily want the



HOUSE OF COMMONS

data shared; and from the individual's perspective you can understand why they might have concerns about some data sharing.

Grahame Morris: We did a major inquiry on deregulated bus services outside London and that was something that we discovered as well. Private bus companies were collecting a lot of data that they were unwilling either to share with the local authority, which was often subsidising and commissioning the services, or to put in the public domain, because of concerns that other operators might steal their most profitable routes. Maybe the answer is to look for another source rather than the operator. Dr Botta, is there anything you could help us with?

Dr Botta: Yes, I entirely agree that the demographic aspect is super-important. That is the first thing I would have said. The second dataset that I believe is missing—I am trying to do some work on that in my research—is on understanding how much it costs to travel. There is a lot of excellent work being done at, for instance, the Urban Big Data Centre in Glasgow, around the time that it takes to travel to places, and accessibility in terms of how many things you can reach in, say, 20 minutes or half an hour, and so on.

It is actually very hard to answer the question when you think about it from a monetary perspective. How far can I travel if I have £50? It is not that easy to answer. It is relatively easy to answer if I ask how many places I can reach in half an hour starting from here, but it is very difficult to answer the question about the £50. Obviously, it is a very important problem to understand how far people can travel in cost as well as time terms. That affects whether they can afford it, which affects their choices. It is a hard question, for which the data is very fragmented. Some of it does not exist, in some areas.

Q15 Jack Brereton: I want to follow on with a little more about the types of data we are collecting, and how we are using it, particularly around investment decisions and Government investment decisions. Often data is used to build business cases and justify particular investment decisions, and how we will capture value from those investment decisions, but there are concerns around how that data accurately reflects the value capture and growth potential.

You mentioned levelling up, Professor Grant-Muller, and that is particularly an issue in areas where we do not have established public transport systems but there is a desire to develop them and develop new infrastructure. How do we use data to make sure that we can build and justify a case, when we do not have the pre-existing data? It is okay in areas like London where there is an established public transport system. We can quite easily justify saying that there is demand and significant growth potential for investing in infrastructure in London. How would we do that in areas in the rest of the country where we do not necessarily have that level of established local public transport? In those sorts of areas how can we build the case for the growth potential and demonstrate how demand could be developed if we were to put in the



infrastructure investment?

Professor Grant-Muller: If I understand the question correctly—

Q16 **Jack Brereton:** I haven't explained it very well.

Professor Grant-Muller: It is very multifaceted. The question is whether data can help us to build a business case for increased transport supply in areas that are a little bit thin at the moment. Is that it?

Jack Brereton: Yes.

Professor Grant-Muller: Well, we collect data on stated preferences and theoretical choices around how people would like to travel and how they currently travel. We can get this, a little bit, directly from people, and of course that already happens. We can also use distributed sensor data that picks up how people are travelling along particular routes at particular times of day, such as the number of private vehicles travelling along particular routes to get to particular places. We can use non-transport type data, such as wi-fi and Bluetooth or particular apps, where we can pick up people's movements, as well. It needs proactive effort to gather that data. It is not necessarily the case that it is lying there to be picked up.

Q17 **Jack Brereton:** Surely that data would just continue to develop the status quo. That is the concern we have. When we think about levelling up, we are trying to change the status quo and see that the imbalances in and around the country are addressed. Surely if we justify investment decisions based on current data and by using it to demonstrate existing demand, we will never invest in any of the communities where levelling up matters.

Professor Grant-Muller: Let me give you an example. If we have the right data, we can understand the maximum time it takes for people in particular geographies to access schools or hospitals, or places of large employment. Where that is above the level that is socially acceptable or at which it becomes burdensome to people, we can see that there is a case for providing better public transport services. If it takes more than 45 minutes or an hour, or over a particular threshold, to reach particular services, and we can see that from some of the data that is there already, that helps to build the business case.

Q18 **Jack Brereton:** Dr Botta, do you have any thoughts on how we use data to justify investment decisions and where investment goes within the country?

Dr Botta: Yes. I mostly echo what Professor Grant-Muller has just said. It is about using existing forms of data, such as surveys that are already being collected, and digital footprint data from mobile phones, apps, wi-fi or Bluetooth, to understand how people travel and move, and where there are bottlenecks in mobility and transport so that it takes too long to reach certain areas, or, going back to my question, where it costs too much to reach certain things like schools, hospitals or even airports. Data



HOUSE OF COMMONS

can help us to understand those bottlenecks and critical points, and to say, "In this area it takes far too long to reach a train station or a hospital at a specific time of day or night." That is where we should invest to improve that situation. In my view, that is how data can help.

Q19 Jack Brereton: Dr McArthur, are there any particular points that you would like to add, particularly around how Government make value for money assessments when they are thinking about investment in infrastructure?

Dr McArthur: As an economist, I should be careful not to trash my discipline, but this is something where economists internationally, and transport planners, do not necessarily have the best record. Forecasts tend to underestimate costs and overestimate benefits wherever they are done, and even the knowledge that things are too optimistic, and correcting for that, still does not seem to fully fix the problem.

Part of it, I think, comes down to data. When we are modelling or making a decision we do the best we can, and where there are gaps we try to make assumptions that we hope are reasonable. The more data we have, and the better it is, the fewer assumptions we need to make, but if you are trying to predict the future, that is always going to be difficult. As I mentioned earlier, transport is connected to everything, so if you are trying to understand some of the large changes, part of it comes down to the models we have, and part of it will be the data that you have available.

Professor Grant-Muller made an excellent point: when we are using data that observes what is happening, that leaves us a bit patchy on what is not happening and who is not moving. Why aren't they moving? What is the reason for that?

Q20 Jack Brereton: And the potential.

Dr McArthur: Exactly. That is difficult to predict from what we observe. You might get other data. Today, someone might get Zoom data. Say someone is in Glasgow and the rest are in London. Why was it a virtual meeting? Maybe the reason is transport-related. Sometimes asking people in stated preference surveys might be the way, but that has its own caveats. If you ask people, yes, they want a train station and they will use it every day. They may say that, if they think it will lead to the funding. There are a lot of studies showing some of the behaviours that you get when you start asking those sorts of questions. Data can be very helpful, but I caution that it probably does not solve all the problems with modelling and forecasting.

Q21 Mick Whitley: I completely understand some of the answers you have given about data in London, but I live on Merseyside, where, for example, the bus and train companies know what the lucrative and not so lucrative routes are. We are battling with companies. We say we want to keep a bus on a route at a particular frequency and they say it is not affordable.



They know that from their data, so they take that bus off. That is the problem for us. They want to be a profitable company, and not just put buses on routes for people on lower incomes who live outside the town and want to go to hospitals and things like that.

It is a difficult situation when we talk about data. The companies know—they have the data—where the profitable and not so profitable routes are. My colleague referred to levelling up. If there is levelling up all over the country, you are talking about billions of pounds, and it ain't going to happen any time soon. How do we square that circle, taking London's transport system as an example?

Dr Botta: It is a very difficult question. There is no way round it; it is a difficult problem. Bus operators and train companies alike want to make a profit. That is effectively their primary goal and it is not necessarily compatible with the fact that in some areas some connections or routes may not be as profitable, but they are still required. It is a very difficult challenge and I am afraid I do not have a solution.

Professor Grant-Muller: I am aware of a particular example, which I won't elaborate in too much detail. A new health facility was built in a place where the cost of providing a public transport connection to it was such that the public transport operators did not want to put on any services. For anybody who did not have a private vehicle it was almost impossible to get to that facility, unless they were willing to pay for a taxi, which is a cost issue. In thinking about the positioning of facilities next to a critical mass of people or where new developments are planned, there is an opportunity for someone to have a look at it in a far more integrated way so that when someone is planning a development they do not only plan the building of infrastructure or a service; they have to plan the transport connections that go with it. A rather stronger steer in terms of the bundle being provided, not just the service but the transport to the service as well, may be one shift in approach that could be taken.

To leapfrog some years ahead, speaking to the people who made that development, what was being mooted was the potential for autonomous vehicles for some of the places where it is difficult to provide services commercially and at a profit. With a future autonomous service the cost model changes, let's say, and that may be the way in which some of our more disconnected communities in future have opportunities to be connected in a cost-effective way.

Q22 **Mick Whitley:** What are the most significant barriers to further developing data usage in the transport sector, and what are the actions necessary to overcome those barriers, Dr McArthur?

Dr McArthur: The transport data strategy does a good job of identifying what a lot of the barriers are. Overcoming them is difficult. Earlier, I spoke about fragmented ownership and the varying incentives of the different data owners who have privacy concerns in some places. One area that we are interested in as academics is skills needs. Some of the



data, particularly the newer forms, may require different computing infrastructure and different skills which local authorities and owners may not be equipped with. Certainly, it is something we are trying to change from a university perspective, but it poses challenges.

There is also the culture around data sharing. The strategy talks about trying to make it open by default. If there is no good reason to hold back the data, it should be shared, but it is difficult for people on the ground making those decisions who often fear that their data might be imperfect or look messy. It probably does, because all data does, so it creates that kind of culture and concern. It is difficult to think about how you tackle that. I'm afraid I am probably one of the academics just highlighting problems rather than good solutions, but at least we are working on some of them.

Q23 **Mick Whitley:** Dr Botta.

Dr Botta: I entirely agree on the skills side, particularly with newer forms of data that require skills that perhaps may not traditionally be the ones people who work in transport have. I do not work in transport; I do not have a background in transport; I have a physics/computer science background. All of the skills that are needed are not necessarily there in the sector yet. As Dr McArthur said, as academics we are trying to do our best to fill that gap.

Related to that, there is a challenge around the people who have those skills. I think it was mentioned in evidence submitted by the RAC Foundation to this inquiry that the people who have the skills do not necessarily end up working in the transport sector. I am stereotyping a bit, but a person who graduates from university having studied artificial intelligence and data science will not work for a bus company; they will try to work for OpenAI or Google where they can do exciting research. That is a challenge that has to be overcome by the transport sector making itself more attractive to the people with the right skills. That can be financially attractive but also attractive in other ways, allowing people to carry out research and other things, because people with the right skills need to go into the sector of their own choice. It is a big challenge. Maybe there are not enough people with the right skills, but the people who have them don't tend to go into the transport sector, in this case.

Q24 **Mick Whitley:** Professor Grant-Muller.

Professor Grant-Muller: I will not duplicate the comments of my colleagues. I agree with them entirely. If we think about certain parts of the transport sector, local authorities and maybe some of the operators and suppliers, I agree that upskilling and increased literacy is needed, but the ambition may be not to have very highly skilled people in all of those organisations, because this is a very dynamic area. To keep those people completely up to date with what is happening and at the forefront of the science would be a huge investment for local authorities, which have lots of calls on their finance and other resources as well.



HOUSE OF COMMONS

I would argue for something a little more nuanced where local authorities and operators have a baseline of literacy and upskilling, so they are educated and skilled, but made aware of where the pools of expertise are and how they can easily access them. That is where the main area of real skill will continue to be. We don't want an Alan Turing fellow in every local authority or bus company, but for those companies to know how they can access that expertise would be a better route forward.

There are some issues about confidence in using the data that is available. That is not necessarily confidence in terms of skill, but confidence in not breaking ethical rules or breaching people's privacy and so on. My experience is that not all the data that is there is being picked up or used as it could be because of an element of fear about overstepping in some of those areas. Finally, we know that things like Innovate UK's calls around innovation in this space are very highly oversubscribed. There is a lot of demand and innovation bubbling up where there are not necessarily the resources to be able to take forward all of the innovation that is there and get it out to industry. That is another factor.

Q25 Mick Whitley: In your opinion, has the Government's transport data strategy effectively identified those barriers, and does it outline the most effective solutions?

Dr Botta: As was mentioned before, I think the barriers and challenges identified in the data strategy are the correct ones. The solution is difficult. A lot of them require multifaceted and very complex solutions. At this stage it is difficult to say whether they are necessarily going to be the right solutions, but in my opinion it is definitely going in the right direction. In particular, the barriers and challenges are pretty much what I would have picked.

Q26 Mick Whitley: Professor Grant-Muller.

Professor Grant-Muller: I agree that the main barriers and challenges have been identified. Some good solutions have been proposed, but this is a foundational document at the moment rather than a very ambitious one or, at this stage, a very highly refined one. Over time it will evolve in that direction.

One thing that I am not entirely convinced of yet is whether all data should be open by default. That needs a little bit more work, because some data could be useful for the transport sector that probably should not be open by default.

Q27 Mick Whitley: Dr McArthur.

Dr McArthur: I think that is true of the open by default thing. I suspect that it would not affect that many datasets in the end because whenever you look at where people are moving it becomes easier to identify them and reveal sensitive information about them. That is tricky and is something we have to think about. The Office for National Statistics talks about its five safes frameworks. Safe data is one of the "safes", but there



are other ways we can make the use of the data safe, by looking at environment settings, researchers and so on. That probably needs to be thought through. If the data can be open, great, but what about other data that we would not want floating around with everyone looking at it?

The other point on the data strategy is that it focused, I think rightly, on the datasets it has access to, or that the Department for Transport would interact with, but there are many other sorts of data sitting around in local authorities and different companies and in other sectors. How that data gets brought together is where a lot of the wins can be, but I can understand that at this stage it is not something they have gone into in depth. I think they mention that it is a living strategy, so this may be something we see developed, and I think that will be positive.

Q28 Chair: You talked about the skills gap in local authorities, operators and others in using the data that is, and could be, available. What will be the role of AI? Is that going to be a further layer of complexity in the skills gap, or is it potentially a help in using that data? Do you have a view on that?

Professor Grant-Muller: My view is that potentially it will be a very useful additional tool, but its usefulness comes with risks. I cannot envisage a situation where all authorities or stakeholders have a full set of AI tools or AI expertise, because it is a very rapidly developing topic. As with the data itself, it is about having literacy around AI, to be able to talk the language and look at it from an informed point of view, but to draw on the huge amounts of resources available in various places—universities and other centres of excellence—and know how to access them.

Q29 Chair: Are there any other views?

Dr Botta: AI is a great opportunity, but I entirely agree that small bus companies will not be able to have an AI expert in-house; it is not feasible or realistic. I think it is about pushing and investing in engagement between the transport operators and the people who have the skills and can support them, and establishing a funding model or some way of encouraging collaboration so that everyone can access the opportunities made available by AI and that will be made available in the future, without the requirement to have an AI expert in-house everywhere, which is just not feasible for smaller companies that will not be able to do that. Thinking about a way of establishing the connection between operators, universities, centres of excellence and centres of research and so on is really important.

Q30 Chair: Dr McArthur, do you have any different points, or would you go along with that?

Dr McArthur: It is a fascinating question, so I am happy to chip in. There are different aspects. AI can provide tools that can allow us to do new things. We have a project using CCTV cameras and AI to develop traffic counts and people counts in different parts of the city. In that case,



HOUSE OF COMMONS

it is a tool where we need experts who are able to design and implement the systems. With some of our students, who might be doing more traditional transport planning applications, you have things like generative AI where AI may be able to help with the construction of a code to analyse data, or it might be able to optimise it. That might take away some of the need for coding skills, but to some extent it is replaced with how you get the generative AI to check your code or optimise it, and how you check that what it is suggesting makes sense. It is interesting, in that it could mean additional skills or different skills. At the moment, it is probably not quite clear what the balance will be, but in 10 years' time this conversation would possibly go quite differently because of AI.

Q31 **Chair:** There might even be some machine sitting here asking the questions for us. Who knows?

Dr McArthur: We hope not.

Q32 **Sara Britcliffe:** The transport data strategy states that the Government's "strategic focus is to enable innovation rather than directly deliver it". Do you agree with that approach, Dr McArthur?

Dr McArthur: That is a very difficult question. I am not sure about that. I certainly think that making the data more easily accessible is a good first step, reducing the barriers to accessing it and using it. Whether that ends up being enough I don't know; it is a difficult question.

Q33 **Sara Britcliffe:** Following on from that, what direct action do you think is needed from the Government?

Dr McArthur: Some of it is tricky. We have talked about skills development and thinking about some of the questions of governance and how the data is owned. I think the strategy mentions things around contracts. When a public body is working with the private sector to deliver something, quite often the contract ends up giving the data to the company or someone else. After the fact, when someone asks for the data and it is not clear who owns it, maybe only then do we realise that the data cannot be used by people who want to use it. I guess there would be a role for Government in thinking about those sorts of governance and contracting and procurement issues.

Q34 **Sara Britcliffe:** Would you like to comment, Dr Botta?

Dr Botta: I would add a couple of things around governance and just thinking about where Government can really do something, perhaps making some datasets available. Things like the Bus Services Act basically made the Bus Open Data Service happen. That is an invaluable resource which has been used a lot. I guess that effectively it forces bus operators to make data available and the Department for Transport is making it available for researchers and for everyone. It is something that has to be thought through and not done in a rush, but making some of the datasets available is really important.



My next point goes back to what we were saying earlier. Establishing models where transport operators, in particular smaller ones, can access the skills that are needed requires funding. It was mentioned earlier that Innovate UK calls are oversubscribed. We have an endless list of unfunded transport projects. Various transport companies have done some interesting—hopefully—innovation around the use of data in transport. Obviously, funding is always limited and I appreciate that it is a difficult thing to do, but that is one of the key mechanisms by which the transport sector can access excellent data that is available from universities, the Alan Turing Institute or more broadly than that. Things that require funding are another area where Government could really help.

Q35 **Sara Britcliffe:** Professor Grant-Muller, would you like to say anything?

Professor Grant-Muller: We have seen a great example of leadership from DFT and other colleagues and stakeholders in pulling together the document. That leadership was very much needed to be able to get this foundational document and strategy. I would like to see that leadership going forward and, as time goes on, having additional confidence and ambition in what the data strategy includes, and to be able to look sideways at other sectors as well.

Data is collected for other reasons. We have talked a little bit about CCTV data, but a lot of that is collected for security reasons. It is often not of the quality that is needed to be able to analyse it for other purposes, for example the transport sector, which it could be. That is an example where dialogue across sectors could increase the value of the data generated and help with some of the business case issues we were discussing earlier. I would like to see that sideways, and cross-sector and cross-Government Department, thinking and leadership, in particular from DFT.

Q36 **Gavin Newlands:** We have gone round the houses a bit on data sharing, but clearly, we are trying to encourage it. Beyond that, Mr McArthur, you mentioned open AI by default. Professor Grant-Muller said there were real risks in that, which you agreed with. I fully accept that, but in moving this forward and looking at what we as a Committee can come up with in recommendations, how can that data sharing be facilitated and regulated to allow for the maximum amount of data to be shared?

Professor Grant-Muller: I'm sorry. I didn't quite catch the last part of the question.

Q37 **Gavin Newlands:** How can data sharing be facilitated and regulated to allow for the maximum amount of data that can be shared?

Professor Grant-Muller: For myself, I would say there is the public good test. Where data is being generated already, perhaps we are not aware of all of it, but it is to ask the question: does that data have a use for public good? If the answer is yes, in those instances where it is needed for the public good there is a very compelling case that it should



be made available, not necessarily publicly available, but there should be access to it under whatever conditions are needed. I would like to see that approach being taken.

Q38 Gavin Newlands: Dr Botta.

Dr Botta: I entirely agree. I think the public good question should definitely be one of the guiding lights for deciding how and when to make data, if not openly available, because it cannot always be, at least available to people who need it.

Perhaps this is going in the wrong direction, as I am raising a challenge rather than providing a solution. Sometimes it is difficult to decide when making some data available would be for the public good or not. In particular, when we are thinking about newer forms of data it is not necessarily clear whether or not using that data would be useful. I think that is a real challenge. Perhaps in the circumstances, things like proof of concept and case studies to see whether or not the data is available and would be useful would be important.

Q39 Gavin Newlands: Mr McArthur, who decides what is the public good? I am sorry to give you a tough question at the end, but you are from the University of Glasgow.

Dr McArthur: Who decides what is the public good? I might come back to that.

On the first part of the question, sometimes it is a question of resource. If you have an open dataset that you can put on your website and people can click on it and download it, that is quite a low-cost thing for an organisation to do. If it is going to require someone to make extracts from the data, or do some kind of processing or set up some sort of secure environment for you to access the data, or go through some application process to make sure it is processed safely, that is the kind of thing that can make collaborations difficult to happen, even if there are no legal barriers and both sides are quite keen on the idea.

Whether a local authority or transport planning partnership has the resource to engage in that and get the data can be problematic. A local authority might have its geographic information science person. They are probably full up with work and then you are asking as an academic, "Could you extract this dataset for me, please?" They often try to help, but may be overloaded. That is part of the issue. We would see more if they had the resourcing available to do it. That can be difficult, because it is not always easy to point to the benefit it would give. If they give me and my colleagues some data and we are working on it, can they point to a concrete output and say it was worth doing it for this reason?

That starts to come to the public interest point. Some of this stuff is speculative. I think the transport data strategy notes that. It is not always clear when data is shared what the outcome will be. It could be quite unexpected; it could generate unexpected benefits and innovation



that can be quite hard to predict ahead of time. Who makes the judgment is difficult. Internationally it varies. Different countries have different levels of trust in different sorts of institutions. Thankfully, academics tend to come off better. When you ask people how they feel about their sensitive data being analysed they often seem quite happy that academics will do beneficial things with it. Sometimes they are less keen on the results, so it is a tricky balancing act.

- Q40 Gavin Newlands:** We have addressed those who are sharing data and potentially using it. Obviously, on the other side of the balance sheet are those whose data we are looking to share. As the issue becomes bigger and more public, maybe we will get a lot more pushback. You do not have to look far; there are lots of conspiracy theories at best about 15-minute cities or 20-minute neighbourhoods, however you want to describe these things. A lot of folk think there is some great big eye of Sauron looking down on all of us and what we are doing. Those self-same people are quite happy to use Citymapper for convenience, but do not realise the link between the two. TfL was referenced earlier. It has been lauded for its open data and how it has been used. What can we learn from the TfL model of data sharing that could be used elsewhere, Dr Botta?

Dr Botta: I am not 100% familiar with the TfL data sharing model. I know a little bit about it. What the public reaction is when data is shared is a very good question, but a difficult one to answer. I may need a minute to think about it.

- Q41 Gavin Newlands:** That's fine.

Professor Grant-Muller: The TfL model of data sharing is well known for having made some of its data very open and accessible, but it has quite a complicated model of data sharing, and quite a bit of its data is not openly available. There is someone who leads on this within the organisation. It has dedicated resource to look at who can access particular kinds of data under what circumstance and for what purpose. They devote a lot of time to assessing issues like what some of the data is going to be used for, where exactly it is going to be stored, what the security arrangements are around the storage of that data, who will have access to the outputs and the model for being able to have a look at outputs before they are made public.

There is quite a hinterland of investment and sophistication to the TfL approach to sharing data, which not every authority would be able to do, I have to say. It has well-developed memorandums of understanding around some of its data and proactively invests in data sharing as one of its policies, which is great. What I have learned from TfL is that some data can easily be made readily available. Other data needs quite a lot of investment and quite a lot of ecosystem to make it available but can eventually be made available.

- Q42 Gavin Newlands:** Have you been able to glean anything from TfL, Mr



McArthur? I suspect there would not be as much data from SPT as there is from TfL, but is there anything you have learnt?

Dr McArthur: Certainly, when cities take a proactive approach, you see a lot more research done on those cities. As academics, you get many more papers published on cities where data is more easily available. It gives them the benefit of much more attention paid to what they are doing and how well it is working, but it takes proactive investment. It is not an easy thing to do.

Q43 Gavin Newlands: Not in any way to speak up for the conspiracy theorists, this is the last question from me, and I will start with you, Mr McArthur, in Glasgow. What are the biggest risks or unintended consequences that we might see associated with developing data usage and data sharing in transport?

Dr McArthur: It is something that we are aware of. Universities, certainly over the past decade or so, have gotten very strict on the usage of data for research purposes. They think very carefully about how it is stored and how it is used to try to make sure that it is used safely and ethically and there is oversight of all of that. It is something that is on our minds. We do not have a right to have data on people and what they are doing. As social scientists at least, we see it as a privilege that we have the data and that we are trusted to work with it. That is something we have to actively work to maintain.

With some of the conspiracy theorists, I am not sure how well the negotiations go or how well the understanding works, but for the general public at least we want to put processes there to keep their data safe, to try to communicate what we are doing and to try to show benefits from what we are doing to the people we have data about.

Gavin Newlands: I am not sure about negotiations. Some of the conspiracy theorists are in government here.

Dr McArthur: I could not possibly comment.

Q44 Gavin Newlands: Professor Grant-Muller.

Professor Grant-Muller: There are potentially, as my colleague mentioned, many unintended consequences and there are opportunities for it to go badly wrong when there aren't proper safeguards in place. It is important that everything that is there is there and that in future maybe there is more as well.

I have observed over the last 10 or 15 years that the narrative has changed. Ten or 15 years ago, we may have been sitting in a room where if we said to people, "There is a tracker on your phone that is saying where you are at particular points in time," there would be looks of horror and gasps around the room, and now very much less so. The narrative is moving on and new narratives are coming along, not all of them helpful, but part of that is to do with public awareness.



In all honesty, the more transparent we are, the easier it becomes. There are some really good examples of authorities—TfL is one—being absolutely up front and saying, “We are collecting your data at this time. You’re being recorded on CCTV or we’re monitoring GPS signals from your phone. Do you want to switch it off?” The more open we are, the easier that becomes and the less opportunities there are for some of the unintended consequences.

Dr Botta: I entirely agree on the openness and the public awareness. I will not repeat that. It is a hugely important issue that will be with us for a few years, as data becomes more available and AI becomes more widespread. There are other issues and other risks. One of the biggest for me is that we have a lot of data but, obviously, we do not have data about everybody. There is an issue of the people for whom we do not have data. They are not represented in the datasets. They are not represented therefore in any kind of AI system that we use to make predictions. Those people are not in the training data, so they will not be represented. Anything that is based on that data will have no knowledge of the fact that those people exist, and that obviously is a huge problem of digital exclusion—excluding people from datasets, either willingly or unwillingly. Sometimes we do not even know who is or is not represented in a dataset. That is a huge challenge that we have and it can go badly wrong if we don’t think about it.

I am always very much in favour of making data open when it can be and I agree that us having the data is often a privilege, but sometimes there is a risk that making data open can be costly, and there is an issue of adoption. There is an issue of making lots of data openly available and then no one uses it and no one needs it, so what was the point of making it available and investing the resources in doing that? It is a relatively small risk, but there is a risk of that happening. The data strategy talks about engaging with stakeholders, academics, companies and anyone who is interested in ensuring that anything that is put out there actually gets adopted, because otherwise you invest a lot of money in making some data nicely available but then no one uses it. There is a risk. It is a relatively small risk, because lots of people want to use data, but it is a risk.

Gavin Newlands: Thanks very much. With apologies to Dr McArthur for calling him Mr McArthur, back to the Chair.

- Q45 **Karl McCartney:** If I could, I am going to ask a general question of all three of you about data, but moving away from the people whose data is being collected to the companies and your access to it as academics. Professor Grant-Muller, you mentioned TfL and perhaps holding them up as an example, but there is good and bad data anywhere. Which companies or organisations are providing good data—TfL might fall into that bracket—but are also keeping some data that they are not letting you see? Can you give us some examples of companies that provide good data or bad data or indifferent data, or do not provide data at all? One of



those companies may be car manufacturers who have an awful lot of data on their vehicles but I am not sure they share it with anybody.

Secondly, do you think that the data that you currently are being given access to will at some point be restricted by cost, and you are going to have to pay for it rather than it being open access? The reason I am asking that is that organisations like the DVLA make an awful lot of money out of the data that they hold. They sell it to various third parties, organisations and companies, and I am pretty sure you do not get access to that for free.

I am sorry to give a huge general question. It is about the data and the access to it now and how you might see that in the future there might be problems, potentially, in having the access that you get at the moment.

I will start with you, Professor Grant-Muller, if that is all right, and then come to Dr McArthur.

Professor Grant-Muller: It is probably worth unpicking a little bit what is good data and what is bad data. What is good data in one set of circumstances or for one set of questions might not suffice for another set of questions. For me, it is whether the data is fit for purpose. The transport sector has been one where historically we have used quite a lot of very murky data, let's say, and we have managed to get quite a lot of out of it. That is probably worth saying.

To step back from that, the way we do research is that we would probably co-design an experiment, a piece of research or an investigation with somebody we think is going to give us the data for it or we would like to give us the data for it. That model of working together right from the start helps to bring forward the data that is needed to answer the questions that we are going to look at. It is very rare that we get into a study of something and then say, "Well, I think we'll have to phone up some car manufacturers," or whatever, and get some data from them at that point. Maybe rightly, we might expect that if we do that the answer could be no, or we might not get the best quality. Working with people from the start is important to get the data.

We have had some absolutely excellent data from FirstGroup. They have been very good at sharing data on various parts of the transport system. They have also been extremely helpful in helping to co-create and to generate data for us—bespoke data for us in some cases—to help us to answer big questions. It is obviously easier where the question is something that they have an interest in and that aligns with their agenda or responsibilities. That is what I would like to say about the good data.

Q46 **Karl McCartney:** Thank you very much. That was FirstGroup getting a gold star from you, but is anybody getting an F, or don't you want to name them? It's all right if you don't want to. I am not trying to tempt you.



Professor Grant-Muller: No, but I think there is a supplier of data that arises from an active wearable that has perhaps not given the data as openly or as wholly as could be useful to us. Let's put it that way.

Q47 **Karl McCartney:** Okay, thank you for that. I will go to Dr McArthur with the same questions if that is okay.

Dr McArthur: The good data/bad data point is an interesting one. As Professor Grant-Muller said, it depends a bit on what you want to do with it. The point I made earlier was that in some sense all data is bad data. Pretty much every dataset will have some kind of issues or problems or idiosyncrasies that probably need to be worked out. Sometimes, the data owners are too sceptical of their own data, or they are too reluctant to share it because they perceive that their data is uniquely bad or has some problems. As someone using the data or working with it, you have to be a bit sceptical about it and have to take time to learn and understand what the dataset is, what the issues are and how they might be mitigated or corrected. When I hear bad data, it makes me think of issues with people not willing to share stuff that could be useful and is probably as good quality as anything else.

Data availability is definitely an issue. I am at the Urban Big Data Centre. Part of what we try to do is to work with people who own data and then make it easier for other organisations and academics to work with that data. We work sometimes with commercial partners and we negotiate about buying the data, paying for it and then have licences to make the data available to other people. It is something that we are exposed to if the business changes their business model or their perception of risk might change.

I used to do quite a lot of work with data from the Strava app. It is typically used for tracking cycling journeys and running journeys. We worked with the company that makes the app and they made data available to us, but over time some of their priorities seemed to shift. You might remember some news coverage of US army bases being revealed because some of the people were tracking their running around the base and then a heatmap showed up where all the roads were on the secret army base. That generated some reluctance to see the data shared.

Covid created more uncertainty. When there were limits on how much exercising you could do and how far you could travel, the company started to get much more nervous about putting out data showing that someone went on a four-hour bike ride when they were only allowed out for an hour. Users might fear that that their data would be used to prosecute them or track them. Sometimes that led to access to data being withdrawn, which can then threaten ongoing research projects or the ability to track things over time if we are studying something over several years.



HOUSE OF COMMONS

It is a real risk that these things change. Trying to work closely with the data owners and trying to collaborate and make sure that it is a win-win situation is probably the best you can do in these situations.

Q48 Karl McCartney: Thank you for that. Is anybody getting a gold star or an F for fail on providing good data from you?

Dr McArthur: I daren't say it, because we will probably end up losing our access to half the data services and data providers.

Karl McCartney: Understood. Thank you for that.

Dr McArthur: They are all wonderful.

Q49 Karl McCartney: Finally, Dr Federico, the same question.

Dr Botta: Thank you for the question. A lot of the things I would have said have already been said. I work a lot with data that is primarily openly available, so I guess a lot of the commercial sensitivities are not present. I do a lot of work with data from the Bus Open Data Service, which I think is great. It is interesting data. It could be better, but it is a good resource for me as an academic. That is a good example.

Things like the Urban Big Data Centre or the CDRC, the Consumer Data Research Centre, are great resources for making commercial data available. At least in my experience, it is not necessarily a case of having access to bad data; it is that sometimes it is difficult to get access to the data at all. I am just an individual academic. For a start, who do I contact? For example, at Strava, as it was mentioned. I haven't tried that, but who do I contact to get access to the data that I might need? That is the first challenge. Even if I know who to contact, why would they give me the data as an individual academic from a university? It is not always a case around whether the data is good or bad. Sometimes the route to access the data is very difficult, and that is the real challenge, particularly for me as an academic.

Some datasets are available commercially. To give you a concrete example, as I said earlier, I am very interested in issues around the cost of travelling and the cost of journeys. The data on the cost of public transport is somewhat available and somewhat patchy, but it is kind of somewhat there. One of the aspects of the cost of travelling is the cost of driving and understanding how much it costs to fill up your car with petrol if you are in Exeter versus London versus Manchester. That question is very hard to answer because, although the data is available commercially, as an academic I cannot afford to buy it at that price. In this case, it was made available to me at a reduced cost, but it was still a cost, which meant I had to include it in a funding proposal, which was not successful, and meant that I do not have the data because I do not have the funding for buying the data. There are challenges with that as well. The way to accessing the data, even before knowing whether it is good or bad, is sometimes the challenge.



Karl McCartney: Thank you.

Q50 Paul Howell: I want to pick up a couple of points you raised, Dr Botta. We have talked about the unintended consequences of data in the transport sector. You touched on it a little bit, but one of the things that concerns me is the missing data, the data you don't have. If we look at what has happened to the economy in the UK, with the gravitational pull of London, everything goes where everything is invested. Does the same happen with data? Because you know more about how travel patterns happen in urban centres, you do more analyses, and therefore it gets better there. Therefore, the rural parts of the country like where Mr Morris and I live see nothing again and everybody just dissipates away. How do you close that gap? How do you as academics get to the spaces where there is less data? Otherwise you have the unintended consequence of continuing to make things worse for people who aren't in the urban centres.

Dr Botta: That is an excellent question. As I said, for me, one of the big challenges and big risks that we have whenever we think about data in general, and in my context digital data and new forms of data, is who this data is representing and who we are missing out. I think of the example of older generations like my parents and grandparents. My grandmother is 107. She does not have a smartphone. If you use smartphones to study mobility, she will not be represented. That is a huge challenge because of exactly what you said; there is a risk of designing systems and frameworks and transport models that only work for the majority of the people who are represented in the dataset. That is a challenge. There are—

Q51 Paul Howell: Sorry, I just want to interrupt and add a little bit to that. Your grandmother doesn't have the phone, so you don't pick up the data. In places like mine where we have villages with people with very low car ownership, very poor public transport, they won't go from a place like Sedgefield to a place like Newton Aycliffe where there are 10,000 jobs because they cannot get there. There should be a demand pattern but there isn't because nobody is going there because they can't get there. I want to cover that piece as well.

Dr Botta: What data and data analytics and data science can do there is to try to identify those areas. It is not just about designing something that, based on the data we have, can make things better, but identifying two areas where we should expect to see many things but we don't, and identifying the things that we do not see in the data and trying to be careful and mindful of that and thinking, "Is that fine? Is that something that we don't have a problem with, or is that something we should try to fix?" Data can really help with that. Yes, it can represent and tell us a lot about the areas where we have a lot of data, but as academics whenever we look at data we also look at what the gaps are. Most of the time, gaps in the data do not tell us anything interesting, but sometimes they do, and that could be one interesting case, looking at scenarios and thinking, "There are lots of jobs here and you should expect to see people



HOUSE OF COMMONS

travelling, but we don't. Why is that? Is there a good reason?" Zooming in and looking into those areas is really important.

More broadly, without going too much into the technical details, there are methods in the statistical community that have been developed to work with datasets that have very poor representation of specific geographical areas or specific segments of the population. There are things like looking at synthetic data and other things so that we try to extract as much information as possible from the little data that we have about specific communities as well. It is a challenge and a risk.

Q52 Paul Howell: You have very largely covered my question, but do either of the other two panellists have anything to add?

Professor Grant-Muller: For new research, when we look for support from the Department for Transport, it now asks for information about inclusion and how people, say, with disabilities are going to be reflected in the research and to what extent we are looking to reach some of the less well-resourced parts of the population. We have to consider it at the point at which research is going in. This started to happen quite recently. I totally applaud, as an initiative, that right from the start we are being asked to be accountable for how the research will accommodate people and geographies who might otherwise be excluded.

Q53 Paul Howell: Okay. Dr McArthur, do you have anything to add? You don't have to.

Dr McArthur: Academics always feel the need to add something.

Paul Howell: A bit like politicians.

Dr McArthur: I think that covers the point. As academics, especially with the types of data that I work with, in many of them there are not enough people represented as you move out of dense urban areas. It tends to get less attention in terms of researching how things are working. It is not to say that the less dense areas are not worthy of development and investment.

Q54 Paul Howell: Okay. Thank you very much, all. Moving subject but still coming back to something that you mentioned earlier, Dr Botta, and that my colleague, Mr Newlands, introduced in terms of the public's concern about data and trying to get some sort of reassurance into them, I understand the difficulty of answering the question about how you get the reassurance, but could you as a group reflect on what particular privacy, ethical or security issues you think people would be concerned about? It is not necessarily how we give them the reassurance but what the actual issues are.

Dr Botta: When we think specifically about transport, one of the key issues is that people are concerned about their location being tracked, as was mentioned earlier. This has changed since a few years ago, but there is still a relatively big concern around being located and being tracked



HOUSE OF COMMONS

and people knowing your location at any given time, knowing where you live, knowing where you go and knowing where you travel. That is a big concern because people do not always feel comfortable with someone else knowing their location at any given time. To me, that is concern No. 1.

Q55 Paul Howell: Okay. Professor Grant-Muller, do you have anything to add?

Professor Grant-Muller: I agree with that. To add to it, data storage has become a little bit more of an issue, not because we have had any particular problems in Leeds or any other university, but because there have been some well-publicised cases of other sectors where data has been accessed, and it creates problems for, in some cases, quite some time for the people whose information has been taken. Data storage is probably No. 2.

Q56 Paul Howell: Okay, thank you. Mr McArthur. Sorry, Professor McArthur.

Dr McArthur: I don't mind. I usually go by David.

Paul Howell: I'll move to that.

Dr McArthur: The point about location data is quite unique. If you give me your location data, it is not obvious to you necessarily what you are revealing about yourself. If you ask me in a questionnaire, "Where do you live and where do you work?", I know that I have revealed that to you, but if I give my location data you can probably tell where I live, where I work, if I visit some healthcare facility and where I shop. That may then be linked to other data, so you might start to be able to identify quite a large number of things about me as an individual just from observing me being at a particular point at a particular time. There is uncertainty about exactly how much you are disclosing and the fact that if the data can be anonymised it is difficult to prove that it is anonymised, and probably if it is anonymised it is not especially useful. That creates some concerns among people about whether they are comfortable with the data being analysed.

The ways to deal with that are things like the five safes framework or the ethics governance procedures that academics work under to make sure that everything has been thought through and that we are only using the data where it needs to be used and for a good purpose, and that it has been handled in a way such that individuals will not be identified. I can certainly understand why there is public concern, and it is certainly something we want to respond to and make clear what we are doing with the data and the fact that we are analysing it in a very aggregate form, and that the conclusions we are drawing are not that someone went to Tesco at this time of day; they are broader and more general.

Paul Howell: Professor Grant-Muller's concern about access is the one that would hit me as an individual. I would be concerned not that I had given permission for somebody to go and do the research but if people



then got that data and tried to use it for other things unethically. That would be my personal reflection.

- Q57 **Grahame Morris:** All of our witnesses have been very complimentary about the DFT's transport data strategy. That is not a bad thing; I'm just making a comment. Is it forward-thinking enough in its approach to emerging technologies? I am asking that in the context of having visited autonomous vehicle facilities in north London where there were self-driving vehicles, use of AI sensors and various other developments. Does this document or policy proposal from the DFT take account of that? Does it address people's perception of risk?

My colleague, Gavin Newlands, touched on this earlier. I remember seeing some survey evidence about what the public think about driverless buses, taxis and so on. They think they should be just as safe as vehicles with drivers in them. The experts, the technical people, say, "Well, they are," but people's perception is that they are not necessarily. This is not just a data privacy issue about location; it is a safety perception. I wonder what you think about that. I will start you, Dr McArthur, if you don't mind.

Dr McArthur: The whole idea about risk and risk perception is a fascinating one. People's perceptions are often quite far removed from what you might be able to measure. Sometimes it comes down to people evaluating risk differently if it is something they feel they are in control of. I seem to remember reading in a paper that, if someone is driving a car, because they feel in control of it they feel it is safer, whereas if someone else is in charge they may perceive it to be less safe. If they are on a train, where they are not in control, they may perceive that to be less safe.

- Q58 **Grahame Morris:** The Committee recently went to Birmingham International Airport where there is a fully automated monorail. I am sure we all felt perfectly safe. It didn't enter anyone's head that it was in any way unsafe. People have a different perception of cars and buses. I have heard the expression "The closer we are to danger, the farther we are from harm." I don't know if that's true.

Dr McArthur: That is interesting. There must be a perception of how complex the task is. We got some driverless trains in Glasgow recently. It is one of the world's simplest networks in the circle it drives around. People seem to be relaxed about that. There is perhaps rightly some scepticism about the task of a vehicle navigating streets that have not been designed for autonomous navigation. I can understand the scepticism. We have yet to see it functioning.

- Q59 **Grahame Morris:** Absolutely. I understand it as well. Dr Botta, do you have any particular thoughts about emerging technologies and public perceptions?

Dr Botta: It is a fascinating question. There is some interesting research, not done by me, on the perception of autonomous technologies. I am not



necessarily an expert on autonomous vehicles. There is some interesting research looking at how we as humans judge the behaviour of machines versus the behaviour of humans. People have done thought experiments where they ask people how they would judge a hypothetical situation when a driver, perhaps to avoid hitting a person crossing the street, ends up hitting more people to try to avoid that person, versus an autonomous vehicle doing the same thing. The answers that we get are quite different. We judge humans for their intention: "He or she tried to avoid a dangerous or difficult situation." We do not judge autonomous machines for their intention but more for the consequences. That is a huge challenge.

Q60 **Grahame Morris:** They have no conscience. That's why.

Dr Botta: Yes. It is a huge challenge because if we want people to adopt autonomous technologies we need people to trust them. The way people trust a human versus the way people trust a machine is very different. We trust them for very different reasons in very different tasks, and that is a huge challenge.

Q61 **Grahame Morris:** Thanks. Very quickly, Professor Grant-Muller.

Professor Grant-Muller: Your original question was what we think about DFT's data strategy and whether we feel that it would span AVs. I have described it as a much-needed foundational document. I don't feel that it is as well developed as it could be, and probably will be in future, to deal with the full range of potential technology advances. AVs are one, but there may be others as well, such as drones and various things. I don't think it is at that level of ambition yet.

There is an expectation, I believe, that when we bring in a new initiative we will plan for it to be safer than the status quo. It is not necessarily the case that that can be achieved. Some innovation will only give an equivalent amount of safety to the one that is there already, and that is a very difficult message to get across because the expectation is of its being better than we are.

Grahame Morris: Absolutely. Thanks very much.

Q62 **Chair:** I would like to conclude the session today by asking quickly about a couple of other areas. The first may be outside your area of expertise, and, if it is, please don't hesitate to say so. It is to do with any concerns that you might have on the actual digital infrastructure that we have, how we collect data and share it. Do we have on trains and bridges enough sensors or the connectivity and wi-fi to make sure that real-time information is shared? Do any of you have any concerns on that practical side of the question?

Professor Grant-Muller: Very briefly, if I may, I have some concerns about sensor supply and the availability of sensors of particular quality. We import quite a few sensors of particular types and are therefore sometimes a little bit prone to what is happening in different countries for



HOUSE OF COMMONS

our supply chain. There are various areas of skill shortage in the ICT sector which are needed and have become costly and very much in demand and very scarce. There are particular areas for my department where we would like to see more of that.

We have a growing set of safe havens for data, and the future investment in those is absolutely key to what is going to happen going forward with data, and the use of data and data security. We have heard about the CDRC and the Urban Big Data Centre, but there are others. The Alan Turing Institute has a data safe haven, as have others. If we were going to do a shopping list of where we need to invest more, there are particular areas where the infrastructure is not quite as complete or as secure as we might want it to be.

Dr Botta: I am not an expert on infrastructure, but the skill side of things is where I see some challenges lying. Even if we have the infrastructure, we need the skills to work with it and analyse the data and so on. That is a challenge.

Infrastructure has to be designed to be resilient. If we rely on a specific type of data collected by a specific infrastructure, what happens if that is not available any more all of a sudden? We have to think about resilience in that sense as well when we design the infrastructure, and think about it holistically as a whole system rather than specifically just one thing, otherwise what happens when that one thing breaks down?

Q63 **Chair:** Thank you. Dr McArthur, do you have any additional points?

Dr McArthur: That more or less covers it. The more infrastructure, the better. I did a paper recently looking at working from home participation and access to fast broadband. It showed that in the areas of slower broadband people were working from home less, which shows that some of the foundational infrastructure around broadband and 5G has impacts on people's behaviour and opportunities.

Q64 **Paul Howell:** Because of what you just said, Professor Grant-Muller—I really don't want to send us down a rabbit hole—when you were talking about international supply chains and centres and things like that, is something that we should be concerned about the potential of those sensors being supplied by places like China, and therefore there being malign actors we don't know about who can get access to the data through the sensors?

Professor Grant-Muller: That is something to be aware of.

Paul Howell: We should leave it there on that one.

Chair: There could be a whole separate inquiry into that.

Paul Howell: As I started thinking about it, I thought this could go the wrong way.



HOUSE OF COMMONS

Q65 **Chair:** We are up against the clock. This is my very last question, and it can have a one-word answer from each of you. We always like to look at what is happening overseas and what lessons we might be able to learn from what other countries are doing. Is there a country that you would identify as the exemplar or the trendsetter in using data for transport planning that we could perhaps do some further studies on? It could be the UK. We could be the head of the pack.

Dr Botta: The UK does very well in sharing transport data and DFT does very well in sharing lots of the data available and collaborating with researchers. It is a very good example to consider. I am not a huge expert on other countries.

Q66 **Chair:** Professor Grant-Muller.

Professor Grant-Muller: I feel reluctant to nominate a country because I am very aware that there are different cultural and ethical backgrounds to the sharing of data that looks ideal in some cases, but we would not necessarily want the rest of it.

Q67 **Chair:** I understand. Dr McArthur, do you have a beacon that we should look at?

Dr McArthur: Norway, because I am familiar with it. They have nice examples of data being made openly available and standardised across the whole country, and some interesting work around linkage to administrative datasets to enrich the transport data.

Chair: Thank you. That is very helpful. We will make some inquiries in that direction. That concludes this afternoon's session. I thank the three of you very much indeed for giving us your time and expertise today. It is very helpful at this early stage in our inquiry to do a high-level examination of the subject. Thank you again. It is very much appreciated.