



## Science and Technology Committee

Oral evidence: [The big data dilemma](#), HC 468

Tuesday 28 October 2015

Ordered by the House of Commons to be published on 28 October 2015.

Written evidence from witnesses:

- [Association of Medical Research Charities](#)
- [Royal College of Physicians](#)
- [techUK](#)
- [Jisc](#)
- [Nesta](#)
- [British Bankers' Association](#)

[Watch the meeting](#)

Members present: Nicola Blackwood (Chair); Victoria Borwick; Stella Creasy; Jim Dowd; Chris Green; Dr Tania Mathias; Carol Monaghan; Graham Stringer; Derek Thomas; Matt Warman

### Questions 1-97

Witnesses: **Aisling Burnand MBE**, Chief Executive Officer, Association of Medical Research Charities, **Professor John Williams**, Director, Health Informatics Unit, Royal College of Physicians, and Dame Fiona Caldicott, National Data Guardian for Health and Care, gave evidence.

**Q1 Chair:** I welcome the panel to the first session of our big data inquiry. We have called it the big data dilemma in order to highlight not only the challenges that utilising the opportunities of big data represents but also some of the risks, which have been highlighted just this week in the TalkTalk data breach. Can I start by raising those very questions? The Wellcome Trust have given us written evidence. They were very clear that “the importance of getting the underlying governance safeguards and communications right”—in terms of healthcare data—“cannot be overstated: the cost of getting it wrong has been evident from the care.data programme...and the consequent loss of trust had ripple effects throughout the health and social care data domain.” Many of the witnesses who have spoken to us have been concerned that the entire debate centres on the negatives of things like data breaches rather than the positives of the opportunities that big data offer, not just to the economy but to key areas of our public sector like healthcare. Dame Fiona, could I start by asking you what you

think we should be doing to ensure we overcome the trust deficit, and move on to utilise the opportunities that big data offer, rather than continually going round the argument about how we overcome the trust deficit?

**Dame Fiona Caldicott:** You used the word “trust”, and that is key to this. One of the absolutely crucial issues we have to think about is how we explain to the public what health data are used for, both in their care but also in the wider interest. To try to do that is quite taxing because the public are made up of many different groups, some of whom are very knowledgeable about these issues, many of whom are not. A characteristic of health information is that many of the public do not think about these issues until they need the NHS, or one of their family does. It is not like banking information. We are often given that example, but many people use it on a very regular basis because that is how they choose to manage their affairs. There are key differences between health data and other data, and I think the task of building trust in the aftermath of care.data is a crucial issue. We need to listen to the public much more than we have. Part of the problem with care.data was that there was not enough listening.

The other point I want to make at this stage is that technology has gone ahead much more rapidly than our conversations with the public. For many people, some of this is quite new, whereas quietly in the background, until publicity like TalkTalk, there has been a huge development of technology, which has a lot of implications for how data are safeguarded and used, but many members of the public have not been exposed to that. I think that is part of how we build trust. Those are a couple of general observations, which for me are important in the context of your study.

**Q2 Chair:** To move on from that, in terms of where the technology is now, to what extent is progress in utilising data dependent on large aggregated and completely anonymised datasets? To what extent is it really dependent on accessing personalised health records? Aisling Burnand, perhaps you could give us some education on that point.

**Aisling Burnand:** I am not a medic, so I will keep it simple. As we begin to understand more about our genetics, we are seeing that in an area like breast cancer you may have 200 variations. People talk a lot about stratified medicines, or what I would call precision medicines—medicines tailored to individuals. I think it is more likely to be cohorts—groups of people. To find those people you will need large datasets perhaps to define precisely the cohort of people you are looking for. Perhaps the best way of thinking about this is how it is experienced at the moment by those with rare diseases. Up and down the country there may be only a handful of people with a particular condition, so the ability to join up public patient datasets to find those people and then use the data for research purposes will, we hope, lead to improvements in treatments and, hopefully, cures and life-saving advances. Based on current understanding, large is going to help medical research, and that is why it is really important.

**Q3 Chair:** The evidence we have received implies that as the datasets become more sophisticated and the capacities to link across them increase, the ability to jigsaw a map between them also increases. Therefore, anonymisation becomes a question and the ethical issues magnify as you go along. That is what was really behind my question about anonymisation and access to personalised records. I am trying to understand what kind of

safeguards need to be put in place to protect the individual while still achieving your aims of delivering more personalised healthcare, and also reassuring the individual that with their data you are achieving your aims of hunting down discoveries. Perhaps Professor Williams might want to pursue that line of thought.

**Professor Williams:** I reiterate that stratified medicine leading to precision medicine is going to be a real game changer. I am a physician, a gastroenterologist. If I could give you a concrete example, we look after patients with inflammatory bowel disease. At the moment when a patient is admitted with a severe attack of ulcerative colitis we go through a series of increasingly dangerous treatments. We start with steroids; then we go for a drug called infliximab, sometimes then to cyclosporine, and, when that fails, at some stage they go to surgery. If we had large datasets, where we could analyse the physical and genetic make-up of the patient and their disease, we would be able to predict, from the large dataset, which of those treatments the patient was most likely to respond to. We would avoid putting the patient through a series of very dangerous treatments and end up precisely with the one most likely to benefit them.

In terms of your question about the risks of big data, we have to balance benefits such as I have described with the risks of identification of the patient, who might come to harm as a result of that. The debate has become confused by the risks to confidentiality of the record at local level with the risks when it is in a very large dataset where, yes, it would be possible to identify the patient. An enthusiastic malevolent hacker will be able to do that, and we will probably never completely eradicate those risks. We just have to get a balance so that the safeguards are there, and we need to explain to the public that there are great benefits but there are some risks, and that has to be weighed up.

**Q4 Chair:** What sort of new healthcare models do you think this big data opportunity will give rise to? You have answered it in part by referring to personalised data and stratified healthcare, but in terms of the actual health service what sort of opportunities do you think it offers in a time of austerity?

**Dame Fiona Caldicott:** Essentially, it is the opportunity to study causes of disease across a large population. A lot of this is already going on in public health. It goes beyond the individual patient to having, as Professor Williams describes, very large numbers of patients who are not identified, and seeing the general outcomes of particular treatments. That has a lot of potential benefit in terms of efficiency, in the sense of not having people exposed to treatments that are not of benefit to them, and understanding how best to deliver a health service that is as efficient as possible. That benefit is really important in a cash-constrained NHS, but the issue is how you take the patient from their individual experience, where they can understand the use of health information for their care, to explaining to them the benefits for the population of having those very large datasets. When you put those questions to them, many people are very content to agree to the second use I described, but they need to have that explanation and the opportunity to consent, with the risks about the security of data that we have all heard about in recent days. Careful explanation and transparency for the public is required so that they trust us with their information so that the beneficial uses can be developed.

**Q5 Stella Creasy:** I want to come back to one element. You are talking about how we might use big datasets in healthcare, and clearly there are some benefits. The challenge for all of us is that we are not operating in a vacuum. When it comes to security and people's confidence in it, the challenge is that, although they might have trust in the way you would use data, data are accessed in the context of other data. For example, you could see a point where somebody could put private data against public data clearly to identify particular groups of people. That is the challenge we face. Even if we could uphold the highest standards of data protection and governance in the public sector on how we might use datasets, we are operating in an environment where other forms of data could be matched alongside it. Have you done much work to analyse what the implications of that might be for data security in the public sector? Somebody could cross-check TalkTalk data against healthcare data, and you could start mapping them together to identify people in a particular locality, for example.

*Aisling Burnand:* We have certainly not done that work; I do not think anybody has done it. It is a challenge. We have to be honest: the way technology is advancing is such that challenges are going to confront us. If we look at the risk side, it is about having an open and honest debate with the public about the risks. From talking to our members, we know that patients want their data to be treated with care. It is their personal data and they want it treated with care. They also want to know that the people handling the data are doing so competently, that they know why they are using it and they are following the rules. There needs to be clarity. For me, clarity is about who is using the data and why, and what sanctions will be in place if people deliberately or maliciously misuse the data, or connect up. At the end, there must be choice. People argue that they need to be able to opt out of something, because they need to balance in their own mind the risk and benefit, and each person will make that decision in a very different way.

**Q6 Stella Creasy:** I think we will come to questions about ownership of data, but I am talking about the capacity to aggregate data, because that would require an additional level of security in how data is used by yourselves. Have you done any work on that? Were there to be an attack on your dataset and somebody got hold of it, would there be the capacity to match it with other data, say open source data, so it is much easier to identify people? My point is that you are not the only ones using data. Have you done much work on building in protections on how you analyse data so that, if it were to be accessed, there would be an additional level of anonymity within it to account for that? If somebody could put together several datasets, which it is entirely possible to do now, what implications would that have for how we manipulate data in the public sector?

*Aisling Burnand:* I am not aware of the jigsawing piece. What I was thinking about when you were asking the question was that, certainly from a medical research perspective, we have done quite a lot of things to try to protect personal data by pseudonymisation, taking away the personal bit and making sure there are safe havens where data can be used. Whether in 10 years' time, in the context of what you are talking about with data, that will be enough I do not know, but certainly from a research perspective a great deal of thinking has gone into how we treat people's data. Something like the CPRD database, which has been in existence since 1987, has led to quite remarkable understandings on the part of the public in terms of making the link between, say, smoking and lung cancer. These are things that have been operating, and there have not been breaches in the use of those data.

That does not mean your question is not valid in the new era, but I am not aware of that work being done.

**Professor Williams:** I am not aware of that. The use of health datasets needs to be very rigorously policed and safeguarded; it should not be put out there to just anybody. The uses of those datasets need to be monitored. If health datasets were then linked to datasets from other sources, it should be with approval by an appropriate body or it is a criminal offence.

**Q7 Stella Creasy:** My point is that there is the potential to do that, which maybe was not there 20 or 30 years ago. Does that have a consequence for the quality of data and your ability to do research on it? If you have to put in additional protections around anonymity, does it have an impact on the ability of the research to offer the results you are talking about? You have to strip out certain amounts of information to make it less likely that it can be used to identify people.

**Aisling Burnand:** To my knowledge, with the systems we have in place at the moment, where we have taken one bit of data away from another—I am thinking about pseudonymised data where we take away the name and address—to date we have managed to handle that safely and securely by having accredited safe havens and places to be able to do the work. I do not want to say that that will always be right for the future, but to date we have systems and processes that have been put in place to ensure that patients' personal data is treated with care.

**Q8 Chair:** We have had a number of submissions calling for a breach of data privacy rules to be a criminal offence, but you have just said that breach of data rules is a criminal offence. We understand that currently it is a civil offence, not a criminal offence. Could you clarify that point?

**Professor Williams:** I do not know.

**Stella Creasy:** It is a civil offence.

**Aisling Burnand:** It is.

**Q9 Chair:** Currently it is a civil offence. I want to follow up on the points Stella Creasy was making. The next issue in this line of questioning is whether poorly implemented attempts to anonymise data lead to further problems in terms of the analysis of the data—whether it will lead to poor decision making or statistical errors in the datasets. We know that we have a lack of digital skills in the UK at the moment, and the question is whether this is affecting the healthcare sector just as much as we know it is affecting the private sector. What efforts are in place, and what work have you done, to assess the quality of the data and the quality of the work being done around anonymisation, especially if there is to be much larger-scale use of big data.

**Professor Williams:** From the Royal College of Physicians perspective, we are concerned about the quality of the source data from hospitals for big data analysis. The process whereby those data are returned centrally is out of date and no longer appropriate for this

purpose. The data are coded by professional coders from patient records that are either paper or electronic. They are coded in statistical classifications—ICD-10 for the diagnosis and OPCS for the procedures—which do not give the clinical depth we now need to answer the sort of questions that are important. What we should be moving towards, and rapidly, is a situation where the data come from electronic patient records and the clinical terms to record the data are coded using an appropriate terminology and coding system. SNOMED CT has been recommended for that. If that were to happen, we would get much richer and more accurate data because clinicians would be validating it in the record. At the moment there is no requirement for a feedback loop for clinicians to validate the data centrally so that we get richer and more accurate data.

**Q10 Chair:** Given that is the quality of the data that is coming out, what kinds of conclusions are being made using that data, and are they safe conclusions?

*Professor Williams:* By and large, because of the volume of data coming out and the limitations in terms of the breadth of the data—just diagnosis, procedures and some demographic data—the conclusions are probably valid, but if we are getting down to much more sophisticated questions, maybe with a smaller population, the conclusions will become suspect. A very topical issue is mortality of patients admitted at weekends. At the moment this is based just on diagnosis. There is no measure of the severity of the illness with which a patient is admitted at the weekend compared with a weekday, so we cannot compare it. It is also 30-day mortality. It is not that patients are dying at the weekend; they are dying over the ensuing 30 days. We have very little information about what happens to them in terms of their treatment and care. In a sense, premature conclusions, not necessarily wrong ones, are being drawn from the data because it is not rich enough.

**Q11 Derek Thomas:** In 2013, NHS England put on hold its care.data system, to which you referred. Can you clarify for us quite briefly why that was, and what you think is the next step forward for the care.data programme?

*Dame Fiona Caldicott:* It was put on hold because there was loud and extensive protest, not least from the general practitioners who were being called upon to download patients' data from their health records about the patient to the HSCIC, in terms that GPs were not content about in relation to their explanation to the patient about the purpose for which that was happening, and what would happen to the data once it went to the central body. I think that reflects the fact that there had not been enough work done with primary care physicians and their support teams and the public about the purpose of care.data. The explanation of the benefits was not clear enough. In the end, given the clamour about those problems, there was an order and it was decided by NHS England that there had to be a pause.

**Q12 Derek Thomas:** Where do you think it will go from there?

*Dame Fiona Caldicott:* The people responsible for the work have been doing a great deal to overcome the concerns of GPs and of patients. You may have heard about the pathfinder practices in four different authorities in the country. They have undertaken to pilot much more carefully worded communication materials with GPs and patients through

the clinical commissioning groups. That work was going on through the early part of this year. It was expected that the communication materials would go out to practices in September, when the Secretary of State for Health decided that new work should be done on the question of patients being able to opt out of how their data were taken from one place to another and used, so there is currently another pause. Were it to be restarted, I think it would be on the lines of much improved communication with both GPs and patients, but a much wider communication programme with the public at large, that this was worth doing.

**Q13 Matt Warman:** Is the problem with the programme or with the communication of it?

*Dame Fiona Caldicott:* That is a very good question. There has certainly been a problem with communication, but there is an opinion that perhaps the programme itself was taking only a rather narrow set of data from primary care to be used to join up centrally with data collected from hospitals. One thing that might be worth considering for the future is whether we should look at a more general question about data flows for a list of purposes, rather than the rather narrow purpose as publicised. It is a matter of opinion, but I think the criticism around that programme is still of an order where if it emerges in its current shape, it may again run into similar criticisms. I do not know that; that is looking a bit ahead, but there is a risk.

**Q14 Matt Warman:** Although if you are arguing to make it broader, the problems we have around getting people to trust, whether it is politicians or doctors, become greater rather than fewer, don't they?

*Dame Fiona Caldicott:* They do, but there is now much more understanding of the kind of communications we need both with the clinical people who are trusted by patients and with the public. I very much hope that lessons have been learned. I think they have. I agree with you that that would be a risk, but something that moved away from the perception that this is the only programme in this area to some of the other purposes my colleagues have talked about could paradoxically be helpful rather than unhelpful.

**Q15 Matt Warman:** You referred to communication by clinical people who are trusted by patients. I think you are right to imply that, on this, clinicians are much more trusted than politicians. Do you think clinicians are willing, able and likely to be making the case with sufficient power to convince patients to agree?

*Dame Fiona Caldicott:* I would hate to speak for all clinicians. There are clinicians who are very dedicated to the benefits to be derived from this and who are willing to do the work. That has been demonstrated by the pathfinders, but there will be an exercise to be done to convince a wider group of clinicians, not just in primary care. You are hearing Professor Williams who is part of the group that is very supportive of this. There would need to be quite a lot of work to get the professions generally on board, but probably general practitioners in particular. They were asked to do something new and politically sensitive and they did not feel they had support and guidance on how to do it. There is a very big communications issue with the profession.

**Q16 Jim Dowd:** Can we move from the role of professionals in this process to that of patients? It is common ground that the whole scheme needs the consent not only of the public in general but specifically of patients.

**Dame Fiona Caldicott:** Absolutely.

**Q17 Jim Dowd:** Can I explore the possible implications of an opt-out for patients? Would that, if widespread enough, affect the value of big data generally and perhaps skew the results?

**Dame Fiona Caldicott:** There is certainly a view that if you have very large numbers of people opting out you lose some of the potential benefits of the large populations we have heard about. As we look at the question of opt-outs, we are learning that there is a hierarchy of agreement among members of the public about what they will readily agree to and what they have more difficulty in agreeing to. If you talk to them about opting-out where their own care is concerned, virtually everybody understands that, if they are to be treated appropriately, their information is crucial to that process. It is when you go beyond direct care to look at things for the health service's use—so-called commissioning—or for research, that you get into a different kind of conversation with members of the public. We have to get much better at doing that, as you have already heard.

**Q18 Jim Dowd:** Is there scope for a differentiated level of consent rather than just the binary yes/no? If so, what would it look like, and would it not be far more complicated to acquire?

**Dame Fiona Caldicott:** Yes. Under the request of the Secretary of State to look at having simple, straightforward communication with the public about opt-outs, we are beginning to understand that the more nuanced and granular it is, the more difficult it is to explain. We may have to make a choice about a very simple question about part of what data are used for rather than all the possible uses, which might then be approached in a different way, but we have not got to that stage in the work yet.

**Q19 Jim Dowd:** The more complicated it becomes, the greater the scope for misunderstanding.

**Dame Fiona Caldicott:** Exactly.

**Q20 Jim Dowd:** Is there any merit in a system that would reward, benefit or incentivise patients to give their consent?

**Dame Fiona Caldicott:** It might be possible to devise one. I am looking to Aisling. Probably her organisation has looked at some aspects of that.

**Q21 Jim Dowd:** You could give them air miles, or something like that.



**Aisling Burnand:** In terms of incentivising, coming back to the risk/benefit piece, it is about making sure that public and patients understand about the systems in place to manage the risks and benefits of the data. What we notice is that the closer a person is to having a condition or disease, the more willing they are to give their data. Some of those with rare diseases are desperate to have a name for their condition. What is really important in this debate is that we are putting the patient first and foremost in terms of their needs, but our views change, as Fiona said, depending on whether you are a well person or you are living with a condition or are sick.

**Professor Williams:** Can I contribute something to the opt-out argument? If very large numbers opted out, clearly that would be a problem, but it would also be a problem if certain groups opted out. If those with mental health disorders decided they did not want their data used, we would have a real problem in understanding mental health disorders and service provision for them and so forth.

**Q22 Jim Dowd:** Do you think there could come a point where, if opt-out was too widespread, you would just abandon the whole thing?

**Dame Fiona Caldicott:** I do not think that at the moment we could abandon it because it is in the NHS constitution that patients have a right to ask to opt out of certain uses of their health data. It may be that they will have to be told in certain circumstances that that cannot be complied with, for instance in an epidemic and certain public health uses, but to withdraw that possibility would require quite a change of thinking. I am not sure that the Secretary of State and the current Ministers are of a mind to go in that direction, but it is possible.

**Q23 Chair:** Do you think that patients would have a greater sense of control of their data in being able to access them? I consider myself to be a net user of the NHS, but I cannot easily access my notes wherever I am. If I could get them on my phone and see what was there, I might feel more comfortable about the full range of NHS researchers seeing them, but I do not even know what is in them. That would make most people feel a little uncertain. Do you think that a core part of gaining public and patient trust is making sure that at least the patient knows what is in their data?

**Dame Fiona Caldicott:** The answer to that is yes. They have been promised that they will have access to their records. I cannot give you an exact date, but it has certainly been published: within the next year or two, access to their records will be available to patients. Whether it will be on their smartphone is another question. Access to the record is one thing; access to all of it on a smartphone is rather different. Yes, I think that would make a difference to patient trust, because at the moment most individuals do not know what is in the record and are not happy that others would see it if they cannot see that it is accurate.

**Aisling Burnand:** It might even help to drive up the quality if they are able to see what is in the record. They might be able to add to the record, at least to say, "Well, that is not our recollection." They would still have to have the health professional involved, but they may even help with the quality of the data. We would certainly welcome greater openness from a patient perspective.

**Professor Williams:** It would also help to drive up the integration of records. At the moment, for a single individual there are disparate records all over the place. It is not just one record in primary care and one record in hospital. There are a whole lot of different records in hospitals as well. We desperately need those to be integrated, and that means standardisation in the way they are collected.

**Dame Fiona Caldicott:** One point I would make in that context is that many members of the public believe that at the moment all those records are available to any clinician they go to see. Part of our difficulty with these conversations, explaining to a patient why they have to go through their history again because they are seeing a different doctor, is that doctors do not have access to all the different records that Professor Williams has just described. There is a real issue for the public about why the ambulance service, for instance, cannot see key aspects of the record when they go to collect an unconscious patient. Some technical issues have to be addressed in terms of joining up different datasets, which comes back to where you started your questions.

**Q24 Chair:** At the moment, pharmacists cannot access patient records either, can they?

**Dame Fiona Caldicott:** Exactly, although arrangements are being made for them to have some of the key details, particularly about the record of medication.

**Q25 Victoria Borwick:** There is quite a lot of enthusiasm on the part of certain members of this Committee to encourage patient ownership of their data. The next generation would certainly expect to have access to it on their iPhone. There is quite a generational change. I sit on a local PALS committee. Even in my own local PALS committee—I know it is not called that any more, but you know the sort of thing I mean—there is a great desire for this. I think most young people would say, “This is my data; I have it on my iPhone.” It gives the opportunities the Chair referred to. It absolutely refers to what you said about accuracy. I was fortunate enough to work on the London Health Commission and met many of the people who are very keen on this system. It helps accuracy and safety, and there is a vast range of benefits to the public, but I do not think any of us round this table has challenged how this is communicated. I am delighted we are having this debate today, and I only hope that everybody suddenly realises why this is important for their own safety as the NHS becomes ever bigger. Can I take you back briefly to some of your Caldicott principles? Do you want to talk a little about what we can do to share information as well as protect it? It overlaps with some of the things we have said already, which I do not want to repeat, but around this horseshoe we are very keen to be able to drive forward the debate.

**Dame Fiona Caldicott:** We were very disappointed when we revisited the new Caldicott principles that we published in 2013, which the Government had accepted, to find that the culture in the NHS of sharing information had not moved in the way we hoped. I think one of the main reasons for that was that we have not done enough on the education and training of staff after they come into post at an early stage in their career. They have a limited induction, if you like, on how to manage information securely and share it, but as they become more senior they are not developing that in the way we had hoped.

One of the really important concepts in the NHS is clinical governance, which is the framework within which we make sure that the treatment is safe, that the outcomes are

registered and audit is done—all the things that ensure we are delivering safe, good care to patients. I have always believed that if information governance had been part of that structure rather than something that was seen as owned by people with technical experience, and actually part of what the clinician does, we would not have had some of the problems we have. I think it is about having that as a more central issue within people's continuing professional development, if they are clinical staff. One of the problems has been that silo effect. Having said that, one of the difficulties we have had over the last couple of years is that, with the disappearance of the strategic health authorities and primary care trusts, a lot of expertise in the technical area has been lost. We do not have a career structure for those who understand how to make the systems you are interested in talk to each other. There is a loss of technical expertise at a time when we need all this to be understood by the clinicians. We need training and education for both groups to have much more attention.

**Q26 Victoria Borwick:** My understanding was that we did own our data but it is, so to speak, kept in trust for us by the Secretary of State. My understanding is that it was going to be released last March, but it was postponed, because basically the clinicians stopped us. I find that regressive, negative and not a positive step forward. Unless we have the debate as to how we can own our data and access it, for all the reasons my colleagues around the table have given, I cannot see the way forward, so I am very keen for you to give us some guidance on how you think we can give advice, remedy the situation and take it forward.

**Dame Fiona Caldicott:** I think there are clinicians who are of a generation—I am not talking about ages—who are of a persuasion that the information belongs to the clinicians. One of the things we have been trying to do is say, “Let's not argue about the record.” The information in the record is the patient's information, and that is what the patient wishes to have access to, wants to be convinced is secure and used in their interests. Perhaps we could leave the issue of ownership. I think that technically the record belongs to the Secretary of State, but that is not the point in this discussion.

**Q27 Graham Stringer:** Can I go back to the early question about the creation of the care.data system? Do Scotland do this better than England?

**Dame Fiona Caldicott:** I believe they do, but I do not know whether my colleagues can assist.

**Q28 Graham Stringer:** Are there lessons to be learned if they do?

**Dame Fiona Caldicott:** The scale is smaller. I think they are able to do some things that are harder to do in England, but perhaps Professor Williams knows more than I do.

**Professor Williams:** They did not have a care.data programme. What they have is much greater professional engagement and involvement in the way data are collected, handled and used. The use of data in Scotland is much better than in England. They are further ahead, because they have put together the infrastructure and the process with patients to do that. It is because of that engagement that they have done better.

**Q29 Graham Stringer:** Changing the subject slightly, the Wellcome Trust and a number of medical research charities are concerned that the European regulations on data protection in this area, which are scheduled to come in next year, will be a barrier to using the data. Do you have any views on that? Are the Wellcome Trust right?

*Professor Williams:* My organisation, the Royal College of Physicians, shares the view of the Wellcome Trust that, if the level of consent required under the proposed European regulation is implemented, it will be impossible effectively to take forward the sort of research that is required, because patients would have to give consent to analyses that probably have not happened because the whole field and the techniques are developing.

*Aisling Burnand:* Like the Wellcome Trust, my members have made that point too. It is a real worry. We hope that the trial or process will come up with the right way forward. When the regulations started out, it was clear that a separate case was made for the research piece, and then it got amended. We are very worried that, if it goes ahead, medical research will be damaged and become unworkable, which will not benefit us at patient/public level; nor does it help all the investment that has been made in this particular area. It will all be for nothing. It is a matter of great concern at the moment.

**Q30 Graham Stringer:** Professor Williams, you talked earlier about getting right the balance between confidentiality and the benefits from the health records. That is obviously the correct position, but how do you put that into legislation?

*Professor Williams:* We need a regulatory framework based on a lot of what Dame Fiona is doing. That framework needs to govern the organisations that have access to the data and identify what would be a malicious or an inappropriate act in terms of access to the data. An ordinary member of the public who hacks into a system because they have IT skills and starts looking at an individual's data by matching up fields would be committing an inappropriate act and there ought to be appropriate sanctions to prevent that.

*Aisling Burnand:* My understanding of the Information Commissioner's Office is that there are already penalties in place that they can impose, but they are civil. Clearly, if we were able to add a criminal penalty, that would be helpful for malicious use. We need to bear in mind that we do not want to make a system that is already very risk averse even more risk averse, but we have to think about doing that through education and understanding what various people are responsible for, so that they understand it better. We must not leave out the education piece if we go down the criminal sanction route.

**Q31 Matt Warman:** This draws together a couple of the previous questions. To what extent do you think that the quality of data we are producing is going to be good enough for the clinical research you are talking about? What changes are needed to make those databases compatible so they work together? On the other side, have you made an assessment of how many lives are likely to be saved—the bare-bones benefits of doing this stuff? In terms of the patient constantly having to repeat their medical history, as you touched on earlier, that is inaccurate and introduces more risks. Have you made an assessment of what cost comes with that, because there must be a clinical outcome cost to that? There are several things, but it is partly about how good the data are and what the cost is of bad data.

**Professor Williams:** There is a report from Volterra which has quantified the savings that could be made through the sorts of things you are describing: the avoidance of repetition, but also the streamlining of care. The potential savings are in the billions. You can argue with their methodology, but there are certainly savings.

**Q32 Matt Warman:** I am aware of the cost savings. I was thinking more of the clinical outcome improvements.

**Professor Williams:** I gave the example of inflammatory bowel disease. I got into the field of informatics because I was frustrated by the number of patients with inflammatory bowel disease I was looking after who came to surgery after a long period of ill health when I had tried all sorts of treatments. After they had surgery, which they were afraid of until they had it, they said, “Why the hell didn’t you recommend that earlier?” We ought to be able to identify which patients are going to respond to which treatments. That would have massive personal cost benefits, and the professions would find it very rewarding as well.

**Q33 Stella Creasy:** You made the extraordinary statement that ownership of the record is technically with the Secretary of State. The alternative to all this is what they have done in Switzerland, isn’t it? There is a Swiss health bank where patients are part of a co-operative. They own their data and opt into what they are prepared to see their data used for, and they get to see how the research is then conducted accordingly. Have you looked at whether it would be possible to bring in a similar co-operative process here? I suspect that most of our constituents think they own their data. Ownership of data will become even more important. Have you considered the Swiss model?

**Dame Fiona Caldicott:** At the moment we are looking at the research that has been done into opt-in systems, but we cannot answer your question as of today.

**Q34 Stella Creasy:** The specifics of the Swiss health bank.

**Dame Fiona Caldicott:** We will do that. We have not done it specifically.

**Chair:** I am sorry, we have a vote. I am adjourning the Committee until we have voted, at which point I am afraid we will have to go straight to the second panel. Thank you very much. You have given us an awful lot to think about. We may come back to you with the final questions we would have asked.

*Sitting suspended for a Division in the House.*

*On resuming—*

## **Examination of Witnesses**

Witnesses: **George Windsor**, Senior Policy Researcher, Creative and Digital Economy, Nesta, **Sue Daley**, Head of Big Data, Cloud and Mobile, techUK, **Dr Paul Feldman**, Chief Executive, Jisc, and **Chirdeep Chhabra**, Senior Manager, Data Programmes, Digital Catapult, gave evidence.

**Q35 Chair:** Thank you so much for your patience. I am very grateful to you for waiting for us. I am afraid we are voting on the Welfare Reform Bill, as you can see. Perhaps we could leap right into some of the evidence we have received and start with some of the benefits of big data. The written evidence we have received from Nesta tells us that “data-driven companies are over 10% more productive than...firms that don’t exploit their data” and that data drives innovation growth in many sectors. Do you think the UK is making the most of these opportunities at the moment, or are there some key proposals that you think we should be taking up?

**George Windsor:** In the submission, we outline some of the benefits, including productivity as you mention, but the stakes for the whole of the UK economy cannot be underestimated. In our sample of firms, we estimate that if the companies that do not exploit their data for decision making were to behave like those that do, it would be associated with an overall 3% uplift in productivity for the whole economy, which is very significant. In previous research we showed that, for instance, the sorts of companies that exploit their data account for about 18% of all companies. It has impacts on profitability, productivity and return on equity. On average, there is a 4.3% higher return on equity, and profitability is about £3,180 higher than average.

In terms of the UK’s position at the moment, a number of great initiatives are going on in the UK, but the landscape is still quite fragmented. A number of the recommendations we made are to try to foster a greater sense of collaboration between the organisations doing these sorts of things in order to avoid duplication and foster a more joined-up system. We made recommendations across schools, universities, the labour market and industry on how some of the skills gaps can be plugged. I can say more about that.

**Q36 Chair:** You have, very cleverly, taken me straight to what was to be my next question. techUK have put in an excellent submission covering the issues. One of the points that jumped out at me was the digital skills gap. You stated that 93% of companies find that the digital skills gap is affecting their commercial operations and recruitment, and it is costing the UK economy about £2 billion a year. You made various recommendations as to how that could be addressed. Do you think this is just part and parcel of the wider STEM skills gap, or is it a specific digital problem that the UK is not addressing in its wider drive to drive up skills and graduates?

**Sue Daley:** Thank you for the opportunity to come and speak to you today. We see the UK as being in a very good leading position to take advantage of big data and the opportunities it presents, but one issue we have raised is the current big data skills gap. You are right. Our submission gives a number of figures. I would like to add that 62% of techUK members have highlighted that there will be a significant gap in big data skills and an increase in big data roles in the next five years. We know there is a problem, and for the UK to continue to remain a leader in this field we need to act now to make sure that we have the right talent pool so that people are there to take advantage of big data jobs; 56,000 new jobs could be created by big data in the next few years.

It is important to remember that, while big data skills are perhaps different from STEM, we have digital skills gaps simply because there is not really one job involved in delivering a big data strategy or solution. Data scientists have a very important role, and that has a lot of focus. The Government's own independent Migration Advisory Committee has highlighted that we have a big data scientist skills gap, but there are many other roles involved in a big data supply chain, or providing a big data solution—from engineers, big data platform developers, data analysts and data scientists all the way up to data visualisation experts. We need to make sure we have the correct skills and people with the right talent at different stages so that organisations can take advantage of the big data opportunities all of us see today and tomorrow.

**Q37 Chair:** You did not quite answer the second bit of my question: are the Government strategies to deliver those tech and digital skills sufficient—are we going to get there—or is there something else they should be doing? Should the Government or industry be doing it? I added the second bit after you gave me time to think about it.

**Sue Daley:** There are two stages, to answer the first part of your question. We need to make sure that we have the right domestic talent pool, so ensuring that we have the right curriculums in place, and a computing curriculum, is very important. We have to make sure we have the right teachers who can teach the right skills that people need, so at domestic level we need to make sure we have the right talent pool. Obviously, Government is working on those aspects with the curriculum issues. The second aspect is to make sure that in the meantime organisations can get the right people with the right talents to fill the jobs they need right now. We need a smart migration policy to make sure that organisations can get the right people to fill the right jobs now. There are areas around the Government's tier 2 visa scheme that need to be addressed to make sure organisations can get the right talent at the right time, particularly now, while we address the domestic pipeline issue.

**Chirdeep Chhabra:** I would like to rewind a little from that. While we talk about the skills gap, there is also opportunity with big data. The volume of data is pretty much doubling every year. Given that situation, we have yet to see enough sharing of data, whether it is personal or organisational data, between different companies—between different silos. That is the element that will create new opportunities and businesses. The skills to fill the gap are there, but we need to start off by enabling sharing of data between organisations. The previous panel talked about the framework of trust that is needed. We are looking at trust frameworks and at the friction in the sharing of data—the legal aspects, compliance and governance, that need to take place to share those datasets. If you do not do that, you will fail in building the next generation of big data companies. Today's very large data companies—the Googles and Facebooks of the world—are all American. They collect huge amounts of data and then exploit it. For us to be able to get into that kind of situation for the next generation of data companies, we need to enable datasharing labs and trust frameworks so we can build an ecosystem around that.

**Q38 Chair:** Dr Feldman, Mr Windsor says we need to have various bits of infrastructure in place; Ms Daley says we need to fix our immigration system and that there is a skills gap; and Mr Chhabra says we need to address the trust deficit and infrastructure. Is there anything else

on the list we need to address in order fully to utilise the big data opportunity, which I think we would agree is out there but there are some barriers in the way?

**Dr Feldman:** My particular focus is on the higher and further education space. We enable a lot of capability that will be coming down the train when it works its way through. We are not coming here to say there is something fundamentally broken in that space, but it is a timing issue; it is a question of people understanding what skills are needed, and then pump-priming to get them through and giving them the experience. We do our best to ensure they have the capabilities to learn when they are there.

**Chair:** Fantastic.

**Q39 Chris Green:** We often hear about the versatility and entrepreneurship of start-up businesses in particular and also of large ones. Dr Feldman, what are the main opportunities presented by big data for the UK economy? Is the UK well placed to take advantage of the opportunities presented?

**Dr Feldman:** It is certainly going to be better in the future than it is today. We are actively working to improve the infrastructure. We connect just about every university in the UK to each other in a whole range of commercial facilities, like Google, Amazon and Microsoft, as well as connecting them outside the UK. Fundamentally, we enable people in universities to access big data wherever it happens to be, along with giving them access to the high-performance computing they will need to process that data. We are also looking at how we can make that available to institutions or companies outside higher and further education. That is our remit. We are looking at how we can provide, in many ways, an income source to universities that have high-performance computing capability, and to SMEs and big business. We have just introduced the first of those with Rolls-Royce, giving them access to high-performance computing based on the backbone network that we provide.

**Q40 Chris Green:** The work going on at universities is contributing to the private sector in quite a big way.

**Dr Feldman:** Absolutely.

**Q41 Chris Green:** Mr Windsor, are there any particular skills gaps in the workforce that you see at the moment?

**George Windsor:** In research we have done recently, one report called “Skills of the Datavores”—datavores are companies that are highly data-driven, with highly data-driven decision making—identified that one of the key skills gaps was getting the right skills mix, having the technical, analytical and industry knowledge to exploit data. That demands excellence in core skills, like maths and stats, and embedding data skills in other disciplines and having hands-on domain experience to understand how to communicate with others in the business. To me, that is fundamentally different from STEM skills and quant skills in a couple of ways. It has multidisciplinary, with all those things meshed into one. It is often referred to as a unicorn if it is one individual, but increasingly



businesses are circumventing the problem of getting individuals who are miraculously endowed with all these skills by building up teams.

The other thing is the location of data scientists in organisations, so that they are connecting the technical and analytical side with the decision makers. They need soft skills and the ability to communicate. It is not just a technician. A data scientist is a real all-rounder who has to be able to communicate findings in a meaningful way to the business. The skills gaps lie in the right mix of skills. We make recommendations about three domains: schools are more about developing excellence in core skills; universities are developing that to a greater extent, embedding it in other subjects, but also seeking high-end data talent; and the labour market is about upskilling workers already in the workforce through CPD and other measures. We make specific recommendations, which I think are in the submission.

**Q42 Chris Green:** With the development of big data, in a few years' time it will be interesting to see whether businesses can develop the right talent so that someone who is very analytical and good with data can progress through a company and get to a very senior position, perhaps even running a FTSE 100 company. Can you think of good examples of people who are in that position at the moment, or who are developing that kind of position?

*George Windsor:* As part of the research, we spoke with a number of different stakeholders across public bodies, but also with some private companies. For instance, we had on our panel Damian Kimmelman who is CEO of DueDil. I guess he is pretty much in that position. I cannot think of any other individuals, but he is very much an entrepreneur in a data-driven company that works with data. That is one good example. It is about embedding it alongside soft skills like communication and creativity, and if businesses empower their workers to play around with data and try to foster innovation within the company. It is also about entrepreneurialism. That is a key part of the soft skills that data scientists are going to need, to drive forward innovation within companies.

**Q43 Chris Green:** Ms Daley, we often think about the lack of STEM skills and our weaknesses in those areas. We have to do a great deal to develop those subjects. Do you think that if we bridge that gap and fulfil our needs in STEM subjects our concerns about having the required skills for big data will be met, or are additional qualities required on top of that?

*Sue Daley:* To go back to what I said previously, having the right talent pool for organisations to find the people they need to fill the roles will be critical. We need to remember that this is not just about graduates; it is also about people in the workforce, so it is about having the opportunity to upskill the workforce to make sure they have the right data skills they need to make the most of the opportunities of big data.

I would take a step further back as well. I think the UK's leadership role and where we are today, which is a very good place, is going to be enhanced and secured by making sure we have a culture of data confidence within organisations, and a recognition of the economic benefits and the driver for productivity, innovation and creativity that big data and data analytics can provide. That really needs to come from within organisations at board level. Board level buy-in and determination to make the most of big data is important, as well as bringing along employees—everybody within the organisation. As we all know, data has

an impact on all of our lives, whether we are at work, on the move or outside. Everybody has a role to play in the way data can help an organisation run, so to build a culture of data confidence within organisations and make sure that staff have the right skills is important both for organisations to get the most out of big data and so that the UK's current leadership is maintained and developed going forward.

**Q44 Chris Green:** It is really positive to hear that we are in a leadership position, despite often thinking that so much more needs to be done. I suppose the big public sector examples of the use of big data will raise that awareness, so that businesses are more familiar with it. Mr Chhabra, do you think the Government have a coherent strategy to develop the UK's big data capability?

**Chirdeep Chhabra:** I think there is much more to be done. There are two aspects to big data. One is about looking at your own data and making sure your business is running better, or you are providing services better, or increasing production, with efficiency savings and so on. That is more internal-looking, but there is also the aspect of mixing data from different sources—silos of data—and that needs to happen more. That is where I think we need more initiatives and what we now call datasharing labs where we experiment with these things. How do we bring together some of these datasets to solve specific problems? That is where the biggest value of data will come from. We need to move more towards that opportunity; that is where the new data companies are going to come from.

There is a whole discussion about personal data. A generation of personal data has already been exploited by the large companies, the Googles of the world, but the next generation will come about through citizens having more access to and control over their data. Younger citizens want more control over their data. If you look at what that will do to the next generation of data businesses, that is where the opportunity lies in terms of the UK trying to make sure it is at the forefront, building labs and frameworks to bring about trust and remove the friction in terms of governance, legal aspects and so on in the sharing of those datasets, so that security is also taken care of and we can talk about successes and are able to prove them.

**Q45 Chris Green:** Perhaps over time we should increasingly see data from health and other elements of government, combined with data from the private sector, being applied increasingly by bodies, public or private, and individuals having access as well, which is a suggestion we have heard several times.

**Chirdeep Chhabra:** Yes.

**Q46 Chris Green:** Is there a great deal more work to do on that, or are we almost there?

**Chirdeep Chhabra:** There is a lot of work to be done—a lot. It is also where the biggest opportunity is, and we need to take a greater leadership position in that.

**Sue Daley:** I suppose that is what makes big data so different.

**Q47 Derek Thomas:** Dr Feldman, you kindly referred to small and medium-size businesses. That is always good to hear. Would you say that the challenges and opportunities available to them through big data are different from those available to bigger business?

*Dr Feldman:* Not necessarily. The issue for SMEs is access to skills and the ability to afford them. Big business just has greater resources to do that, but in many ways SMEs will be a bit more fleet of foot, and possibly a bit more creative, in the way they use data. That is why we are trying to see how we can help SMEs to use university facilities a bit more, where they are free. There is nothing fundamentally different; it is more about whether they can get the capabilities they need to the level they need.

**Q48 Derek Thomas:** Is it a question of the cost attached to even accessing big data—whether they can afford it?

*Dr Feldman:* So much depends on the nature of the big data. A lot of it is open data, increasingly so, where they are as able to get at it as big business. The data big business has that is private to them is the data they collect. Every second, Tesco collect an awful lot of data on us that they do not share with anybody; they have a great wealth of data, whereas an SME can have only the private data they can access or buy, and the buy is always a bit suspect. The availability of open data is where they can really make a difference in that sense.

**Q49 Derek Thomas:** Are we seeing much collaboration between smaller businesses and big businesses like Tesco?

*Dr Feldman:* I would not say collaboration; it is more about whether they can get contracts to help. That is where a lot of SMEs are fleet of foot and creative, and that is how they grow in many ways. I am sure people around me here have more to add on that.

**Q50 Stella Creasy:** The example of the mydata project answers a lot of the questions you have raised. That was an attempt to create standard portable and accessible information. Although it is my data, I cannot access anything meaningful from my energy company to share with other energy companies. That hampers consumers in accessing data, and it also hampers citizens in relation to any of the benefits you are talking about. Clearly, there are commercial disadvantages in sharing data because people can compete for your customers. It also means we can create new industries and jobs for people. How do we square those circles? Do you think that we as citizens need rights to our own data, to be able not just to take it from one company to another, but to share it en masse with a third party, say a new company, if the benefits of big data that you are talking about are to accrue to the public?

*Dr Feldman:* I am not sure I am qualified to answer that question.

**Q51 Stella Creasy:** Were you involved in the mydata project at all?

*Dr Feldman:* No, I was not.

**Chirdeep Chhabra:** I completely agree that there is a movement towards citizens being given more and more control over their data. If you look at the general data protection Act that is being discussed in the EU, it is interesting that that is the direction even the younger citizens of the world want. They want to have more control over their data. If that is so, we will end up in a situation where basically we have control and can give access for specific reasons to specific organisations. That brings about a new level playing field whereby you do not have organisations just scraping data off you. The reality is that in the doubling of data every two years, two thirds of that data is personal, and that is going to grow; there will be a lot of internet of things data and smart city data. A lot of it will be personalised.

**Q52 Stella Creasy:** There is data that is commercially useful and which will drive a lot of these industries. That is why companies do not want to share it. Obviously, it puts them at risk of loss. As consumers we do not have access to that data in any meaningful fashion. Do you think it will require legislation? For example, in Australia there is legislation on this. If I were an Australian citizen, I could request my data and share it with a third party; for example, a third party could negotiate a new service or industry on my behalf with an existing company, which we cannot do in this country right now. Do you think that is where we need to get to? Otherwise all the benefits from sharing that you are talking about—

**Chair:** We are running out of time. Can we have some quick answers, please?

**Chirdeep Chhabra:** It will definitely happen. It is bound to happen. That is our view.

**Q53 Stella Creasy:** Have you seen it happen in any other country without legislation?

**Chirdeep Chhabra:** Not yet.

**Q54 Chair:** The statement was just made that companies do not want to share data because it is commercially sensitive. Is that your experience? Is there a desire to share data, or is there a blanket desire not to share it? Do you think that is an accurate statement? Could you comment on that briefly before we move on?

**Chirdeep Chhabra:** That is an accurate statement. Some of the things we are doing are to create safe havens where organisations can bring their datasets together. They are not giving data to each other; of course, they will not do that for governance and business reasons, but organisations are realising more and more that they can only get benefit from their data by mixing it with other data outside to improve their supply chains. They are dependent on other industries too. There are no mechanisms today. There is a lot of friction in datasharing, in terms of legal governance and so on. That is where we need to take a lead. If we do, we will have many more such industries coming along.

**Q55 Carol Monaghan:** Dr Feldman, you talked about companies like supermarkets collecting data from customers. It is well known that that can be done with reward cards, and then marketing is targeted as a result. Are supermarkets at the moment collecting data from customers who are not using reward cards? Is that something of which we need to be aware?

**Dr Feldman:** I have no way of knowing that. Sorry.

**Q56 Jim Dowd:** The answer to that question is that they certainly collect data from credit card purchases. They know people's spending patterns that way, not just from loyalty cards. Anyway, I will be brief; I am aware of the Chair's exhortation about shortage of time. We discussed medical research earlier. What other areas of research do you think would benefit from big data? What other opportunities are there?

*Sue Daley:* There are opportunities for big data across all different sectors: media, finance and in particular sport, which is a key area for data analytics. A lot of sporting companies are working with academics to look at that. In terms of science, the Hadron collider creates about 16 million gigabytes of data a year. That is a huge amount of data. Data analytics and big data tools are helping scientists to unlock the mysteries of the universe. Perhaps that is useful to the Committee. That is another area where I definitely see big data in action right now.

**Q57 Jim Dowd:** Are UK universities well placed, and do they have the facilities, to exploit that?

*Sue Daley:* We have seen significant Government investment, which techUK has supported, in working with academics, not just in setting up the Turing Institute; we are working with the Harwell Centre in bringing industry and academia together in academic research. That has had a boost from the big data investment the Government brought, so we are well placed. The question now is what we do with that. How do we apply the insights and knowledge that can be gained from that kind of work to the wider society?

**Q58 Jim Dowd:** In your experience, are the universities and research institutes adapting their teaching methods to encourage students to understand the opportunities provided by big data?

*Sue Daley:* It is probably for others to answer that. What we have found works really well is where industry and academia come together and work together to engage and collaborate early in the process, so that students can see the big data opportunities out there throughout their careers, but it is probably for others to answer that question.

**Q59 Jim Dowd:** Does anybody else want to answer that?

*Dr Feldman:* I could come back on a couple of the questions you asked. Universities are getting a lot of the investment they need; for example, we will connect universities to the Hadron collider and the Harwell Centre.

**Q60 Jim Dowd:** Are they getting enough?

*Dr Feldman:* Can you ever get enough in that space? They would always say they want more. A lot of money is going into the space. The key thing is that you have facilities like the Hadron collider, and a key part is that the universities wherever they reside need to be able to access those. The UK has certainly invested to ensure that that happens. That is the key advantage we have over most countries, and we should glory in it, in that sense.

**Q61 Matt Warman:** Coming back to skills, how can we ensure that young people are empowered partly to make sure they understand that this is their data and they should have a right to see it, but also so that they have the enthusiasm to want to work with it, from schools right up to universities?

**Sue Daley:** We have a great opportunity to bring understanding that data has an impact on our lives, on every aspect of our lives, and the younger generation will know that more than any other generation ever. We need to address it in schools and help build a culture of data confidence in schools and then build it up from there so that young people understand the role of their data and gain confidence in how to use it. There is also a good opportunity to rebalance the public debate and talk about the good uses of data and how data positively impacts on all our daily or digital lives. The more we can raise awareness and perhaps promote the good uses of data, and how important it is in terms of driving our digital economy, the more we will encourage young people—from my point of view, particularly girls—to get involved in this industry and the more we will build a UK society that understands the power of data and what it can do and, therefore, is part of the data journey we are all on.

**Dr Feldman:** As an example of how two of these things come together, we are piloting with universities a concept called learning analytics, which is taking the data we can capture about students at university, making it available to them and looking at patterns. There are examples, both in the UK and elsewhere, of being able to predict from that the success or otherwise of students, and to plan interventions. We imagine that over time it will go into their school records and back through their lives. That has to be made available to students. They will understand the data from a comparatively younger age than perhaps they think about it now, and will see the impact of data on their lives and learning. We need to help them understand how that can be of positive value to them to help them succeed.

**Q62 Matt Warman:** Does this tap into the point Mr Chhabra made earlier about the inevitability of data opening up if we can embed it into curricula and younger people's lives?

**Dr Feldman:** Absolutely. If you embed it into people's lives, suddenly we all understand the pros and cons of data and become much more savvy about what data means to us, and what we share and protect.

**Chirdeep Chhabra:** We are data-driven anyway. The apps we all use, even to travel on the London underground, crunch huge amounts of data and there will be a huge amount of productive analytics from them too. It is about bringing to the fore exactly how everybody is using that.

**George Windsor:** One particular recommendation we made was around embedding data analytics in other subjects. One really good example of where this is happening is the Urban Data School in Milton Keynes. They are using local data. Obviously, it works in Milton Keynes because it has a lot of sensor data—transport data; it is a smart city. They are embedding data that is meaningful to students within other disciplines. It might be for geography, so they look at meteorological data that is meaningful to them. They know that last week it rained heavily. Why is that? They explore and visualise those data. That is one

way of trying to embed a sense of passion in young people about working with data. Coupled with that, it is providing better information about career prospects in this field at a young age. The Tech Partnership is doing great work on this. For instance, their tech futures programme includes tech careers, tech for girls and those sorts of things. We need that to be scaled up, and it needs to be more specific to data.

**Q63 Matt Warman:** Briefly, what should Government be doing to make that happen more?

**Sue Daley:** I would point briefly to the very good example of Nesta to make it clear that not all big data is personal data. Obviously, weather data is not personal data. Not all the data analytics involved in big data are to do with personal data. It is important to remember that there are different types of data. I am happy to provide the Committee with more details on that, if it would like them.

As to what the Government need to do, we need a joined-up approach and to make sure that we address the skills gap and have the right skills talent. We need to rebalance, and have the right public debate about big data so that we can move forward and, by doing so, address the trust and confidence issues we have about data. We have a great opportunity and the UK is in a brilliant position, but we need a future vision. Looking forward, one of our recommendations was to have a long-term strategy as to where the UK is going. How can we build on where we are now and make sure that our future is just as bright as today in terms of big data and the opportunities?

**Dr Feldman:** One of the ways we hope to do that is through a data skills taskforce. One of the recommendations in our report “Analytic Britain” was for a cross-cutting taskforce, and we are going to convene a meeting on 11 November, which will include the likes of techUK, the Digital Catapult, the Tech Partnership, the UKCES and others to discuss some of the skills issues and where we can try to plug skills gaps.

**Q64 Chair:** We started this session by questioning how we close the trust deficit. I have a statistic from techUK’s submission, which said, “cyber-attacks cost UK businesses £34 billion per year”. We have had TalkTalk this week. I also have Digital Catapult’s “Trust in Personal Data” review. I was struck by one particular statistic, which is that in terms of dataholders the public have most trust in the Government and banks. They are at 40%ish, and everybody else—all private sector holders—is down under 5%. I find that extraordinary given that we blame the banks for the financial crisis and the Government is full of politicians who are the least trusted people in the world, apparently. What is it that Government and banks are doing that is different from the rest of the private sector and leads to that huge disparity in trust levels in data collection?

**Chirdeep Chhabra:** When we look at what the Government and maybe the banks do with our data, the element we do not hear about is banks losing our data every now and then. We know that retailers collect a lot of personal data, but they do not tell us what they do with it, so there is no transparency around what is done with that data. Although we know that the cards people use in Tesco and so on collect data, we have no access to any of that. The same thing applies to NHS data, which the last panel talked about. While we know that the data is about us, we have no access to it. Those are some of the elements that are in question. There is a trust deficit, and for companies like Tesco and other retailers it is

much higher because they are using our data, but at the same time there is no transparency or access control over how—

**Q65 Chair:** Is there more transparency with what the NHS is doing with our data?

**Chirdeep Chhabra:** Yes. But you do not hear about the NHS losing the data every now and then.

**Q66 Chair:** Care.data caused a huge drama. We have had HMRC scandals and all sorts of others. There have been lots of scandals, but for some reason there is much more trust—a dramatic difference: a 35% difference—in public sector holders of data and banks as holders of data. Have you done any research as to why there is that specific gap, because that is something we would want to try to understand quite urgently?

**Chirdeep Chhabra:** I will make sure I get some answers back on that. My colleague was leading that effort. I will make sure we report back on that.

**Chair:** That would be super. The report was my weekend reading. I was fascinated. Thank you very much, panel. You have given us a lot to think about. I hope you feel that as our inquiry continues we are contributing to an important national drive to make the most of big data opportunities.

### Examination of Witnesses

Witnesses: **Richard Woolhouse**, Chief Economist and Head of Research and Strategy, British Bankers Association, **James Meekings**, Co-Founder and UK Managing Director, Funding Circle, and **Imran Gulamhuseinwala**, Partner, Ernst & Young, gave evidence.

**Q67 Chair:** Welcome, and thank you for joining us. We have invited you here as a FinTech panel, particularly because the Government have identified FinTech and the development of commercialisation of new financial business models and disruptive innovation as a key area for growth. We think that intersects very significantly with the big data inquiry. I do not know whether you were listening to the end of the evidence we received from Digital Catapult, which was that, even though there is a huge public trust deficit in the gathering of data, for some reason the public trust the public sector and banks far more than they trust all other private sector organisations in terms of data collection, to the order of 40% trust in the public sector and banks and under 5% for all other data collectors. Do any of you have an idea why that might be?

**Richard Woolhouse:** Maybe it is the long history of collecting data. Banks have much more experience and have been doing it for many more decades than perhaps other sectors. I don't know; I have never seen any research on it.

**James Meekings:** Arguably, there is greater regulation of financial services than some other parts of the private sector and that helps to give them confidence and creates blueprints for companies to gather and store data—how they use it and transfer it and so



on. Maybe that pushes through to how consumers understand that their data will be used. As a young company in financial services, I know that our customers have lots of questions about how their data are used and stored, which is probably different from older financial services. If you ran the same survey with us, probably people would be less than 40% more trusting. It is probably different segments within the market.

**Q68 Chair:** That is interesting.

**Imran Gulamhuseinwala:** I am not able to speak about the level of trust outside financial services or the FinTech sector. One of the things we notice when we speak to customers of our large clients is that they are very much behaviourally in a mode where, if something goes wrong with the banking relationship, the bank will make good financially. We think they take that mindset across from the financial make-whole to the data make-whole. There is not a great deal of evidence to suggest why that would be the case, but it creates a level of comfort in the mind of the underlying consumer.

On James's point, I would be surprised if the FinTech community, which is much more about newer companies that do not have the same level of state support, had that level of trust. It can take a very long time to build up. One of the things we spend a lot of time thinking about is the level of adoption among the FinTech community, and confidence around data is one of the issues raised.

**Q69 Chair:** Given there is this trust deficit and it is playing a significant role in big data and its success, and in the sharing of data between companies as well, to what extent will big data be a factor in the expansion and success of the UK FinTech sector?

**Imran Gulamhuseinwala:** If I take a step back and think about the role of data in the FinTech space, our research analysis has found that it is absolutely critical. There are three or four enablers to FinTech. We do not need to spend time here talking about the supply side enablers, but on the demand side the key area around adoption seems to be the ability for consumers conveniently to share their data with some of the FinTechs and be confident that their data will be looked after in a trustworthy way. If you like, we can spend a bit of time talking about the lots of different ways in which FinTechs and banks can utilise data, but the fundamental issue about whether, if the data is sent across, it is a trustworthy source is one of the key determinants of adoption.

**James Meekings:** Funding Circle is a technology platform which connects investors and businesses over the internet. Our role is to provide that platform but also to help understand the risk of different businesses so investors understand what the expected losses would be from those loans. The role of data in that process is for us to understand the risk and help inform the human underwriters who look at those loans. It is critical for our business, because the better we get at it and the more data we have to do it, the fairer the price is for small businesses, and therefore we can help small businesses. The competitive edge is huge as well, because it allows us to get more customers, build better datasets, because we have more data, and spend more on marketing. As a reinforcing cycle, data enables players to become better and better, which you probably do not get in any other industry. Our whole product is a technology platform that prices risk, which is enabled by data.

**Q70 Chair:** As you build those bigger and bigger datasets which interact with each other and you understand the risks more and more, how does the consumer understand the risks more and more?

**James Meekings:** We believe in complete transparency. Unlike a bank, on Funding Circle you can download the whole loan book. You can see every single loan we have done; you can see its status, whether it is repaying or not and so on. As we get more data, we publish it to the market. That allows every customer to have scrutiny of us; obviously, not every customer does that, but a level of people do. In financial services, when companies have data there is a big question about how much of it should be transparent and released to people. That is not to say everyone is fully informed and understands it, but it allows scrutiny of the industry and therefore makes sure that practices are right. As we develop, our commitment is to carry on publishing that data, making sure we show people who are putting their money in that this is our expected loss rate, which we predict, and this is what we actually have achieved. We can only do that by being fully transparent with data.

**Q71 Chair:** How common is that practice among your FinTech peers?

**James Meekings:** I can only speak for the lending industry because that is where we are based. We have a trade body called the Peer-to-Peer Finance Association, which sits on top of formal regulation by the FCA, and in that we have certain criteria for membership. One of those criteria is: publish all your data for everyone to see. We think that is incredibly important. In an industry where it is not us but the investors taking the credit risk, that exchange mechanism can work only if we give as much data as we can to those investors.

**Q72 Dr Mathias:** Can you tell us a little more about Funding Circle and the partnership with British Business Bank Investments and how that came about?

**James Meekings:** Two or three years ago we engaged with Government on a number of requests. One of those was helping Funding Circle grow to allow us to take more institutional capital on board. One way of our doing that was to have all of us in a room with the Government as taxpayers lending through Funding Circle. To date, the Government have committed about £60 million and they are earning 7% on it, so we should all be relatively happy with that today. It has been a good thing for industry because Business Bank has to do due diligence on the process, understand how we do credit and so on. It has been very positive.

**Q73 Dr Mathias:** Has the Financial Conduct Authority made any difference to your business?

**James Meekings:** Yes. For everyone in the room who does not know, peer-to-peer lending is getting regulated at the moment. We have just submitted our authorisation. It is the right time for this to happen in the industry. We have always wanted it to become regulated. In excess of £1 billion is going through these platforms on a yearly basis, so it has to be regulated.

**Q74 Jim Dowd:** Perhaps Mr Meekings is best able to answer this. Is there a difference in the way new SMEs and start-ups use big data compared with the more traditional financial institutions?

*James Meekings:* Different companies use different datasets, but speaking for ourselves, we use the same underlying data as banks. We use data from Experian, Equifax and credit bureaux and mix that up with the data we request from our businesses when they apply for a loan. The underlying data is pretty much the same. It is what we do with it. Within a bank, banks use data in different ways for different types of customers. In the consumer world, banks are very automated and are quite advanced in how they use data. On the business side, they are not. We are using techniques that banks use but we are using them for customer services that banks do not use them for, if that makes sense. In a way, what we do is tried and tested, but it is being applied to a new customer set, allowing us to help more small businesses across the country.

**Q75 Jim Dowd:** Does that provide a better customer experience, or does it just fit your economic model?

*James Meekings:* In my slightly biased view, I think it does provide a much better experience. The most important thing for small businesses is being able to get on with running their business. They want to be able to run things quickly. Being able to apply for a loan online, speak to someone over the phone and get it done within a matter of hours is a much better experience than going through a two-month process with a bank.

**Q76 Jim Dowd:** Do peer-to-peer lending networks have the potential to disrupt the traditional operations of the banks, or is that your objective?

*James Meekings:* I certainly think we do. I think we are complementary at the same time. We have many partnerships with banks. There are certain types of loans we can do because of our risk appetite, or our investors' risk appetite, that banks would not want to do. We are competitors on one side and complementers on another.

**Q77 Jim Dowd:** In our first session—I do not know whether you were here—we looked at medical data and people's sensitivity about that. Surely, the same thing applies to people's personal financial information. Some would say it is more important to them than their medical history. Are there implications for the FinTech industry in that?

*James Meekings:* There could be. Speaking for Funding Circle, we believe that if people are lending their own money they are entitled to understand the risk. We cannot pass on to investors every bit of information we take. For example, we cannot pass on to investors company director information because it is personal, but everything that we can pass on we believe we should, because ultimately it is people's money that is being lent.

**Q78 Jim Dowd:** Mr Woolhouse, do you have any observations on those questions?

**Richard Woolhouse:** I very much agree with James's point that there are complementarities between the banks and some of the alternative lenders. They have referral arrangements. It is a segment of the market that is growing very quickly; it is growing at about 100% a year and it is about £1 billion. It is still small compared with the aggregate stock of SME loans in the economy, which is a couple of hundred billion pounds, but it is providing a complementary function too.

The other point to note is that we have seen a slight shift in demand patterns in terms of SME lending space. There is an expansion of net lending to mid-size firms, whereas for the S-s it is a bit flatter in terms of the net lending profile. That probably reflects the fact that there are other sometimes complementary or alternative providers who are able to fill the gap, given how the post-crisis regulatory landscape has changed the economics of some of those businesses.

**Imran Gulamhuseinwala:** One of your questions was about how banks view data versus how the FinTechs do. A reflection from my organisation would be that the banks are not using the data in anywhere near as creative a manner and from as many alternative sources as the broader FinTech community. I am talking about the broader FinTech community rather than James's specific segment. To give you an example, historically the banks organised their businesses around products, so they find it quite hard to move data between the different product silos. You probably have some familiarity with it when your bank does not know your credit card information on the mortgage side and vice versa, whereas the FinTechs have very much started with a clean piece of paper and are using information data from as many different sources as they possibly can. They are using unstructured and social media data, sometimes to help with refining the product proposition, but often to help with client segmentation or client acquisition, and sometimes to improve the overall client journey. Part of the reason they have done that is that many of them thought of themselves as technology companies at the outset. The other part is that there is data asymmetry between the FinTech community and the incumbent banking community, and that often lies around the current account information at both the consumer and the SME level. It is very much the case that the FinTechs do not have access to that data, so they are trying to create a picture from a variety of different sources of data, of which there has been an explosion through mobile social media and so on.

**Q79 Chair:** In your report, you say that the UK has a leadership position in peer-to-peer lending, and that we are strong in emergent FinTech but poorly represented in traditional FinTech. Why is there that separation?

**Imran Gulamhuseinwala:** There are a couple of pieces to that. The UK benefits enormously from some underlying factors that have proved very attractive to new FinTech—emergent FinTech. On the demand side, it is the quality of the underlying market. You have consumers with high levels of connectivity, penetration of smartphones and so on, high wealth and high adoption of non-financial services: for example, online products, e-commerce and so on. And of course you have the City of London. A combination of those two means it is a very interesting market to be in. When you add to that the fact that the regulator, by which I mean the FCA, has been rather helpful in fostering innovation within the FinTech community, through initiatives such as Project Innovate, you have a very interesting combination of factors. London itself is also an

interesting talent hub, where you have both technology and financial services in the same place, unlike the US where they are split on either side of the country, and increasingly there is the willingness of private capital, particularly VC, to come across.

**Q80 Stella Creasy:** I have two topics or questions that I want to put to all three of you. We talked a bit at the beginning about the high trust in the banking industry. Given that you are more likely in your life to be a victim of fraud, if not crime, online than you are offline, do you think that is merited?

**Richard Woolhouse:** I do not know. That is a difficult question.

**Q81 Stella Creasy:** One of the challenges we have talked about in the Committee today and previously is that at the moment data breaches are a civil penalty. There would be quite a reputational risk for banks were there to be a persistent level of fraud. Therefore, there is a commercial advantage in reporting it, as opposed to addressing the concerns people might have or the financial consequences for people. Do you think we need to change the way we do data protection in this country to create a more secure and honest environment when there are data breaches?

**Richard Woolhouse:** A lot of co-operation occurs both between banks and between the banks and the authorities around cyber-attacks and financial crime. There is an obvious balance to be struck between the ability to protect an individual's data privacy and the ability of banks effectively to monitor financial crime by sharing data and talking to each other. The concerns about some of the European directives are that that will be constrained, but an enormous amount of work is done between the banks and between the banks and the authorities. We have seen recently that, in order to secure banking data, that needs to happen across all sectors of the economy, so it is not just a specific issue about safety in the banking system.

**Q82 Stella Creasy:** The flip side of that is the mydata project. The one place where they made progress was around financial services and the standardisation of data. I want to ask you, Imran, about the client journey. As we get to a point where there is sharing of data that allows me to move my bank account and have confidence in the banking system, is there also a case for moving all the data that come with that account? If that is what banks are going to make decisions on in future, my ability to access a good deal will be influenced not just by the data you hold formally on my account but by the other information on which you may have based a decision about whether to lend to me. What are the implications for consumers on that basis?

**Imran Gulamhuseinwala:** To start with the first question, I am wandering into the realms of *Papal Economics* and so on, where you get the sense that a new social contract needs to be constructed around data that does not sit well with our typical categorisation of industry sectors. Yes, I can see the situation that you are more likely to suffer from fraud online. When I look around—I do not have the data to back this up—the majority of it is happening in non-financial services in online marketplaces, online billing and the secure custody of passwords and credit card information.

**Q83 Stella Creasy:** The reality is that it is everywhere, but the third question I come to is whether we should have more honest reporting about when there have been data breaches and what data has been breached, because banks are creating bigger datasets. What rights do you have as a consumer to that information?

**Imran Gulamhuseinwala:** I agree that there needs to be almost a new taxonomy for consumers to understand this. I feel quite strongly that Government has a role to play in terms of consumer and SME awareness. A couple of things are happening that make this, in our view, inevitable. One is PSD2—the payment services directive—which puts payment initiation very much in the hands of third-party payment providers, so there will be another level of potential intermediary that may not be a recognised financial services brand, and actually may not even be a financial services regulated entity in the traditional sense of a deposit-regulated entity. It will be sitting in and among the ecosystem, so consumers will find it very hard to understand which particular entity is responsible for what at any point in time. Very soon you lose the opportunity for them to say, “I trust financial services but I do not trust e-commerce.” It is all blending together, and that is one of our big observations. One of the things FinTech is doing is forcing industry conversion across all those different verticals, and allowing people like Uber, frankly, to offer payment services without having any bank associated with it at any point in the customer journey.

For the benefit of the Committee, I should point out that I am on the steering committee of the open bank working group which is addressing some of these issues from a specific banking consumer and SME point of view. It is far too early to talk about any of the findings of that particular group. Mydata has been talked about within that context. There is a broad feeling that mydata has been a very interesting, robust first step in enabling consumers to understand that they have transaction level data; it belongs to them and they can also use it for their own benefit. Mydata has a very narrow use case which is about trying to shop around for the best current account. None the less, it feels that it is beginning to move in that direction. Adoption of mydata is very low—a point you referenced—largely because, it is fair to say, it is incredibly inconvenient.

**Richard Woolhouse:** The other day nearly every other transaction was blacked out as well. It may have had something to do with who owns it.

**Q84 Stella Creasy:** Given where we are going with data and the huge opportunities that come with it, but also given the issues around security and lack of clarity about whether it is a civil or criminal issue when things do go wrong, and given the difficulties for the consumer—I was very struck, Imran, when you talked about the client journey and how banks are using social media; I am not sure my residents would know that perhaps their social media postings were being used by their banks, or even supermarkets, to make decisions about selling to them—do you think there are rights or particular powers consumers themselves should have to drive a new market in these industries? James, this may also be a question for you; I don’t know if you are cutting across that in terms of your customers. Would it also offer them a protection that at the moment they do not have, either because people do not understand what they can do with their data or because the company, when you start to negotiate with them,

does not offer you the data but finds ways round it because it is not commercially advantageous to them to share this kind of information with people?

*James Meekings:* My view is that data help us get more small businesses cheaper loans.

**Q85 Stella Creasy:** Sure. But should your customers have a right to request data from you in a particular format that is portable and transferable to a third party across the board? That would include all the data, and they would also be told, for example, if you are going to look at their Twitter feeds when you make decisions around loans or mortgages, for example.

*James Meekings:* In my view, yes, because it is about transparency and who owns that information. I think people should be able to use that information. In your example of people using mydata to switch credit cards or current accounts, you have a perverse situation where the people who have high costs and most need to switch are probably in overdraft, and by switching without any data they get the worst rates. You have a perverse situation where no one uses it because of that downward spiral.

*Richard Woolhouse:* Current accounts are not a great example because you are switching between products that are essentially pretty similar and free, with some exceptions. They are not actually free, but they appear to be. It is different from an energy bill where you will save £80 or £100.

**Q86 Stella Creasy:** But for your industry it has some commercial advantage in that you can compete for other companies' business, but there is a commercial disadvantage to having datasharing, which is one of the reasons why mydata has not worked. The lesson for us in the broader use of big data is that, if we want all these benefits, what rights do we need to give to citizens or consumers that would drive some of this, because there is not necessarily as much industry interest in it as we might think?

*Richard Woolhouse:* In the SME space a lot of data sharing goes on. Credit scoring affects decisions.

*James Meekings:* Yes.

*Richard Woolhouse:* But those are all voluntary arrangements between different participants in the marketplace.

**Q87 Stella Creasy:** Credit scoring is a great example. I know we are going to come to that. It is why Wonga and all those companies did not do credit sharing. It was not in their interests to feed into that industry, and as a consequence banks were lending to people who were in debt.

*James Meekings:* It becomes a very interesting challenge. You have credit bureaux where anyone can access the data, and then you have other companies sourcing lots of different bits of data because they currently cannot get it from the banks. Once they have done all that clever work, does it become their IP? The underlying data is not theirs, but the process of getting it is. It is a really complex issue.

**Chair:** Thankfully, that leads us perfectly to Derek Thomas and the final area.

**Q88 Derek Thomas:** We are going to finish with new ethical challenges—a light-hearted subject. Would you say there are ethical challenges for the financial technology community, as opposed to the long-standing traditional financial services?

*James Meekings:* Ultimately, we all sit within financial services, so I would not expect there to be any difference. If we are gathering data and using it, we should be treated in the same way as banks, and banks should be treated in the same way as us. I cannot imagine there would be a different ethical stance between newer players and traditional ones.

**Q89 Derek Thomas:** In terms of how people access the services you provide, it is slightly different from going into a bank. For example, if I was a vulnerable person, maybe someone with a learning disability, would I have the same access to your services, and would you have greater ethical challenges, as opposed to if I just went into a high street bank or some other more traditional service?

*James Meekings:* It is true that there are fewer ways to access us. We still have a phone-based system and online as well. I do not see a significant challenge that I am concerned about at the moment. I will probably have to give the question a bit more thought.

*Imran Gulamhuseinwala:* Could I ask you to repeat the question?

**Q90 Derek Thomas:** Basically, what we are asking is whether, in terms of ethical challenge, there are new ethical challenges that present themselves to your sector or your community—the financial technology community—that might not exist for standard traditional banks that may be situated in the high street.

*James Meekings:* You can see us all struggling.

*Imran Gulamhuseinwala:* The question posed earlier is probably as close as I have seen to that. For example, iwoca provides working capital finance to businesses—sorry, that's another competitor of yours—and uses a lot of revenue data from, say, Amazon. They pull that down, but they use it to provide a service, effectively, lending in real time. You could make the argument that if they start using social media or Twitter feeds and find something disparaging about an owner, perhaps they will begin to factor that into their thinking. I do not know that the banks are not doing that, frankly, themselves; employers are. There is certain information in the form of data in the public domain already that has not been initiated by a financial institution. It is out there, and the financial institution, FinTech or incumbent, may choose to use that as part of their overall algorithm for writing safer, more profitable business.

**Q91 Derek Thomas:** In terms of access to social media and my online existence and whether or not that makes me creditworthy, do we need greater legislation? Do we need some clarity? Would an individual, or a younger person, even know that that might be looked at in terms of whether they are creditworthy?



**Imran Gulamhuseinwala:** On the social contract point, consumer awareness needs to be understood, so that when you put things on those kinds of sites that is the public domain, at least as far as this Committee views that information. There is an incredibly significant practical issue, which is that because we do not have an aggregated sense of identity it is incredibly hard to know. If, say, you came to me as an incumbent financial institution and I decided to look at your Twitter handle to see the kind of stuff you post, I might find 15 different Derek Thomases out there. I am not sure which one it is, but because I am trying to run very quickly and am processing thousands and thousands of things I just take the top five, for example. You can imagine doing something like that and trying to get meaning out of it. If you are not one of the top five, that would be an unfair representation of your data, but I do not see that as an ethical situation that anyone has tried to create; I see it as a problem of identity within the UK.

**Q92 Derek Thomas:** It is not necessarily unique to your sector.

**Imran Gulamhuseinwala:** Absolutely.

**James Meekings:** For most companies that do it, when you log into Kabbage or iwoca, you are creating a social contract at that point, so the question would be whether people are aware of that when they start logging in to get a loan. You would think they would be. In relation to Twitter, there are 20,000 businesses registered on our site, so there is no way we would know what their Twitter handles are, because they are just random things.

**Chair:** I think we are getting a big hung up on Twitter and the social media point. What is behind this, as Stella Creasy said, is that Wonga is an example of a company which used algorithmic credit ratings and turned round very fast judgments, and we have all seen the headlines about the consequences that followed. I think it is disingenuous to say this is not a new problem and that it is not a different environment from traditional banks where a longer process is followed, often with face-to-face interviews. What I am trying to understand is what processes are put in place to try to monitor and address the obvious ethical challenges that arise when you are using an algorithm to assess somebody's life.

**Q93 Stella Creasy:** I do not think it is so much the algorithm that is the challenge as the source data. You have the capacity to gather large amounts of information in a way that 20 or 30 years ago you did not. Should a consumer know or have a right to withhold it and say, "You cannot make a judgment about my eligibility for a mortgage based on what I tweet"? That is coming, and it is something investment companies and others do. We talked about Tesco earlier. Tesco collects information from your Facebook and uses that to market and target; Google does it. The information—the potential—is there. What we are trying to get at is should we, as we progress, have laws or protection for consumers from the consequences of that, because there is a disparity in understanding that it is happening?

**James Meekings:** My view is that it has to be clear and transparent to the consumer when they are signing up to these sites what it could be used for. Then they can make the decision at that point, as opposed to a law saying they cannot do it.

**Q94 Stella Creasy:** Do you sell on data?

*James Meekings:* No.

*Imran Gulamhuseinwala:* Certainly FinTech businesses do that. Monetisation of data is a core pillar of some of the economic models in some situations. What FinTechs tend to be at the moment are very monoline product providers; they tend to get a lot of business that they then need to reject because they cannot deal with it. That is absolutely fine; it happens in lots of different online models, but they can resell those leads and package them up. Yes, that can be part of it.

**Q95 Stella Creasy:** Do you think that protection needs to come in? I take your point about consumers: buyer beware that you are buying a product. What we are surprised about is the extent to which this can be done. It is done for marketing purposes at the moment, but it is not without plausibility that it could be used at a more granular level to make better decisions. At what point do we step in and say that, for example, consumers must be informed that their data are being sold on, or that they must be informed that their social media profile is being used as part of the decision-making process? We do not have those laws at the moment because it is not something people have ever dealt with before.

*Imran Gulamhuseinwala:* I think it makes sense.

**Q96 Matt Warman:** In a practical sense, is it remotely possible that we could ever enforce a world where you—plural—are not allowed to use publicly available data? It seems pie in the sky to me.

*James Meekings:* It is a question of whether we are not allowed to use it or whether it is clear to the customer when it does get used.

**Q97 Matt Warman:** Obviously, it is in your interest to be transparent, but it does not matter whether we think you should not be doing it; you are not going to stop.

*James Meekings:* That is why you would have to think about the source of the data. Is it behind a wall, or is there some way of dealing with it out there? You might have some nice people like us who say we are not going to do it, but some would.

*Imran Gulamhuseinwala:* For example, one of the things Wonga did that was a key determinant in their algorithm was to time how long it took someone to fill in an application form. If you did it too quickly it probably meant you were looking at three or four different sites; if you did it too slowly, it meant you were too unsure. Would you say that is a third-party piece of information? I make that point just to show how creative some businesses can be in how they think about data. It is so fast moving that it will be very hard to create a long list of public data repositories that people use. There will always be the next thing. For example, telematics in the insurance industry is all about, frankly, a contract about sharing data on driver behaviour, timings and usage. That is another classic example. That will be augmented with other information: for example, road signs. I can see that you are driving at 40, but were you doing it in a 60 mph limit or a 30 mph limit? That is the third-party piece you are sucking in. The list could go on and on.

**Chair:** We could probably discuss this all evening, but I am afraid we have come to the end of the session. Thank you so much for the evidence you have given. It has been really helpful and interesting, and we will probably write to you for specific technical detail as we go along. That brings this session to an end. Thank you very much.