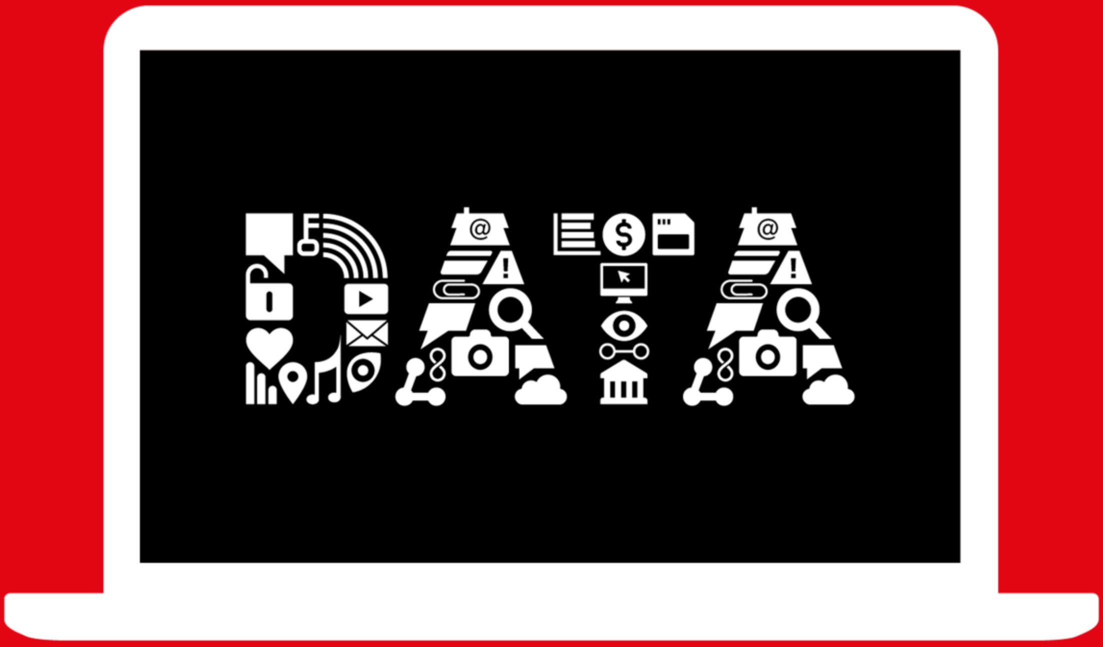
June 2020
APPG AI Evidence Meeting

BIG INNOVATION CENTRE

Data Governance

Beyond GDPR

PARLIAMENTARY BRIEF



***Data Governance: Beyond GDPR** is a* Parliamentary Brief based upon the All-Party Parliamentary Group on Artificial Intelligence (APPG AI) Evidence Meeting held online on the 24th February 2020.

This Evidence Meeting was chaired by **Stephen Metcalfe MP** and **Lord Clement-Jones CBE**.

We would like to express our appreciation to the following people for their oral evidence:

- **Dr Florian Ostmann,** Policy Theme Lead and Policy Fellow, The Alan Turing Institute
- **Dr Kenan Direk,** Research Data Manager, Institute of Health Informatics UCL
- **Tamara Quinn**, Partner - Intellectual Property, Data and AI, Osborne Clarke LLP
- **Ellis Parry**, Data ethics adviser, Information Commissioner's Office
- **Andrew Pakes**, Research Director, Prospect
- **James Kingston**, Deputy Director, HAT LAB
- **Professor Ryan Abbott**, Professor of Law and Health Sciences, University of Surrey

Big Innovation Centre is the appointed Secretariat for APPG AI

- CEO, **Professor Birgitte Andersen**
- Rapporteur: **Dr Désirée Remmert**

The video recording of the Evidence Meeting can be found on our websites.

www.biginnovationcentre.com | Email: info@biginnovationcentre.com | @BigInnovCentre
www.appg-ai.org | Email: appg@biginnovationcentre.com | @APPG_AI

PARLIAMENTARY BRIEF

Data Governance: Beyond GDPR

All Party Parliamentary Group on
Artificial Intelligence

APPG AI Sponsors

The group's supporters – Blue Prism, British Standards Institution, Capita, CMS Cameron McKenna, Creative England, Deloitte, Dufrain, Megger Group, Microsoft, Omni, Osborne Clarke, PwC, Rialto and Visa – enable us to raise the ambition of what we can achieve.

Contents

1. Introduction

The aim of this APPG AI meeting was to discuss current issues around data governance, specifically concerning the GDPR, in the UK. Specifically, we addressed problems emerging with the collection of personal data, its commercialisation, and its use by third parties. In this context we also discussed alternative ways to govern and share data in the future.

aippg
appg
All Party Parliamentary Group on
Artificial Intelligence
EVIDENCE MEETING
DATA GOVERNANCE: BEYOND GDPR
FEBRUARY 24 AT 5:30PM
HOUSE OF LORDS
EVIDENCE GIVERS
DR FLORIAN OSTMANN
POLICY THEME LEAD, THE ALAN TURING INSTITUTE
TAMARA QUINN
INTELLECTUAL PROPERTY AND DATA PRIVACY PARTNER, OSBORNE CLARKE
ANDREW PAKES
RESEARCH DIRECTOR, PROSPECT
ELLIS PARRY
DATA ETHICS ADVISER, INFORMATION COMMISSIONER'S OFFICE
PROF RYAN ABBOTT
LECTURER IN LAW
UNIVERSITY OF SURREY
JAMES KINGSTON
DEPUTY DIRECTOR
HAT LAB
DR KENAN DIREK
RESEARCH DATA MANAGER
UCL INSTITUTE OF HEALTH INFORMATICS

The APPG AI Evidence Meeting convened a group of experts from academia, think tanks, and business:

- **Dr Florian Ostmann,** Policy Theme Lead and Policy Fellow, The Alan Turing Institute
- **Dr Kenan Direk,** Research Data Manager, Institute of Health Informatics UCL
- **Tamara Quinn**, Partner - Intellectual Property, Data and AI, Osborne Clarke LLP
- **Ellis Parry**, Data ethics adviser, Information Commissioner’s Office
- **Andrew Pakes**, Research Director, Prospect
- **James Kingston**, Deputy Director, HAT LAB
- **Professor Ryan Abbott**, Professor of Law and Health Sciences, University of Surrey

This meeting was chaired by **Stephen Metcalfe MP** and **Lord Clement-Jones CBE.**

Parliament has appointed Big Innovation Centre as the **Secretariat of the APPG AI**, led by **Professor Birgitte Andersen (CEO)**. The Project Manager and Rapporteur for the APPG AI is **Dr Désirée Remmert**.

This brief will first present the evidence given in Parliament by our expert speakers as well as additional suggestions by a member of the advisory board. Our speakers addressed the following questions in their evidence:

1. **What are the various models for data ownership?**
2. **Should we think of data as a commodity or as a common good that can deliver public value?**
3. **Can data be democratised without harm for the individual?**
4. **What conclusions can we draw from the application of the GDPR so far?**

The brief will conclude with a summary of the main arguments and the resulting policy recommendations at the end of the brief.

2. Evidence

Dr Florian Ostmann, Policy Theme Lead and Policy Fellow, The Alan Turing Institute

I would like to focus on three issues in my evidence. The Alan Turing Institute is the UK's national institute for AI and data science. Our headquarter is in London, but we form part of a network of about 500 researchers across the country located at different universities. I will address this topic partly from the perspective of a researcher, drawing on the challenges that researchers are presented with in their work. However, I am also a member of the Public Policy Programme at the Alan Turing Institute. Our group focuses on two areas. One of these research areas is concerned with **governance challenges that arise in the context of data driven technologies.** We are working together with the ICO and the FCA. The other area deals with the with the **use of data and the realisation of benefits of data use in the public sector in the pursuit of social good**. I will use examples from these two areas to illustrate my point.

First, I would like to address the issue of **consent in the context of collecting and processing personal data**. Specifically, I would like to discuss the issue of **meaningful and informed consent**. During the implementation of the GDPR there has been a huge increase of attention to the processes that ensure formal consent – consumers ticking boxes and clicking “ok” to confirm they agree to the “terms & conditions”. However, there is still much left to do to **reliably ensure that consent is informed and meaningful**. We all know from experience that we agree to these conditions without having read the terms and their details. Areas in which this is particularly important are those where **the very same data can be used for beneficial purposes but also harmful purposes**. One example from financial services is the use of data around vulnerability on the part of firms. It could be used in a way to protect consumers, for example in overcoming certain health issues they face. However, the very same data could also potentially be harmful as it could applied in a way it would negatively affect consumers’ credit ratings. In order to enable the beneficial use of that type of data is very important **to provide assurance to consumers that data is only used in a beneficial way**. That is a challenge in terms of ensuring good governance, but also when it comes to communicating the use of these terms to consumers. The current method of providing consumers with **multiple pages of terms and conditions is not effective** in this regard. We need to think of **best practices to communicate this in a precise way**.

The second issue I would like to address are **privacy enhancing technologies in the context of data sharing**. Data trusts have already been mentioned. The APPG AI has looked at this and related topics already in the past. Those ideas concern the organisational prerequisites of sharing data. Yet, there is also a technological angle to this. This concerns mostly the **removal of technological barriers to data sharing**. One example concerns privacy enhancing technologies such a **differential privacy or federated machine learning**. These methods can be very effective in addressing some of the concerns that are in the way of data sharing. However, there are two challenges:

1) **All of these technologies are in their infancy**. The lack of best practices, established standards, and general knowledge of the abilities and limitations of these technologies are slowing down their implementation.

2) There is a **lack of clarity from a regulatory point of view** in terms of what the implications of these technologies are – especially when it comes to the **distinction between personal and anonymised data**, the **risks involved in deploying** these technologies, and how they should be addressed from the **regulatory compliance** perspective.

The third point I wanted to share concerns the **distribution of value when it comes to the use of personal data**. Personal data governance debates tend to be **focused on control and consent**. These are important issues, however, another important question is often overlooked: **Who benefits from the use of personal data, especially in economic terms?**

One way in which this is often discussed is in terms of the question if consumers should be paid for the use of their data. However, this is just one angle one might take. There are other questions to raise, for example about competition – the increasing recognition of the fact that depending on the level of competition in the context of a service economy, where services are for free and revenue are generated through data processing, **the same service might be provided to a lesser or greater degree of privacy interference**. It is important to bear **the value dimension of the exchange in mind**. Of course, there are also questions around patent protection or profits derived from products developed based on personal data or publicly available data and how that value is shared in the economy.

Dr Kenan Direk, Research Data Manager, Institute of Health Informatics UCL

I would like to start by thanking the organisers for their invitation to give my perspective on data governance in the academic research space. I currently work as a health data manager at the University College London Institute of Health Informatics, whereby my interests are centred on the use of national electronic health records (EHR) of 35+ million people to provide real world evidence that positively impacts public health.

Recent exemplar studies include (1) The first and currently only **chronological map for over 300 physical and mental conditions**, offering unprecedented insights into the most fundamental epidemiological question: **who gets what and when?**; (2) **100 million primary care consultations indicate clinical workload may be approaching saturation**, prompting questions on **how the nature of the patient-GP interaction must evolve** to contend with future demands; and (3) currently underway and the first of its kind, an ambitious programme of work that will use **routinely collected data to better understand the health of over one million migrants in England**.

These studies represent, in my opinion, the very best of what we can do with the health data of millions of people. But how are we able to do this? **Have millions of people consented to their data being used for research? No is the short answer.** Under the General Data

Protection Regulation (GDPR) framework, the lawful basis for this type of activity is outlined in Articles 6(1.e) and 9(2.i and 2.j), which collectively describe **a legitimate public interest in the processing of non-consented data for scientific/public health research**.

Within the remit of public health research, we may be interested in potential drug side effects that did not emerge during clinical trials or if a drug developed for one indication can be repositioned/repurposed for another indication? Therefore it may not be surprising to learn that **your health data is sold to academic and for-profit organisations in order to perform this type of work**. Unfortunately we have lost transparency through these data transactions and that breaks the social license.

The true potential of individual health data at population scale can only be realised through genuine engagement between the public and organisations that can translate data into products of therapeutic value. We have the tools to conduct health data research securely (e.g. data safe havens) but we do not bring the patient with us through this journey. What could the health research environment look like in 5 or 10 years? It is not impossible to **conceive a landscape where data never has to leave the health service and benefits both the patient and the researcher -** the patient can see what data is recorded, who is using it, and for what purpose; whilst the researcher has access to the world largest and richest medical resource underpinned by the most advanced e-infrastructure and data governance models.

Tamara Quinn, Partner - Intellectual Property, Data and AI, Osborne Clarke LLP

I am a partner at Osborne Clarke. We are a law firm headquartered in London with a network of overseas offices. I specialise in IP and data and am part of the AI and Machine Learning team. I come with the of view of a humble practitioner advising clients on very practical, real world day-to-day questions around the development and use of systems involving AI and machine learning.

I am going to talk about two things today. **Interactive property** and questions around **ownership of data used for machine learning systems**. I will also touch upon the **GDPR** and some of the issues arising from its application to date.

Intellectual property protection:

The **data sets upon which the systems can be trained** are increasingly recognised as **important assets**. As was mentioned in the **European Commission's Data Strategy Paper** that was published last week alongside the **White Paper on AI**, one of the concerns which is slowing business take-up of AI is **uncertainty over what IP rights there are in data itself**. It is worth understanding that most companies which generate or collect data are not AI companies. They are not technology companies and they certainly do not even see themselves in the data business in many cases. That is where a lot of our potentially valuable data is sitting. In collections of data, in particular **non-personal data**, which are **generated as**

a more or less unnoticed side effect of day-to-day business operations of these companies**.** Think of the so-called **"digital exhaust**" - information that is bulged out by our increasingly connected environment.

When we talk about data ownership, what IP protection is there for data? The reality is, that it is patchy. It is **different from IP protection for AI and machine learning systems** themselves, where there can be **quite good protection in terms of copyright and patents**. This is quite **different for data**. What we are forced to rely on is a patchwork of database rights, available confidential information, and trade secrets. We obviously try and bolster that with contractual obligations where we can, but I think it is worth mentioning that **database rights, despite their name, are actually limited in scope**. To qualify for protection, without going into all the details, the creator of the database has got to make substantial investment in obtaining, presenting, or verifying the data in the database. This will cover things like spending money on already existing data, seeking it out, compiling it, verifying it, cleaning it up, and so on. What it does not generally apply to is investment in creating data in the first place. This means that the sort of data I was talking about earlier will often just not be covered by the database right. I think this is potentially an issue. It means we ought to at least look again at the question of whether we ought to be making any **changes or modifications to IP law to broaden the protection** available.

When it comes to datasets, I am very cautious about pronouncing as to whether we should have included protection. However, I think it is something that we should be thinking about. However, in any case we should create **clarity for businesses** so that they can spend their money on going about their business rather than delightfully spending it on lawyers trying to work out protection.

GDPR:

Moving on to the application of the GDPR to date and what conclusions we can draw. We have had quite **a lot of guidance from the ICO and the European Data Protection Board**. Not so much in terms of formal enforcement of three of the courts, or through the ICA. There obviously have been some high-profile enforcements around data breaches. However, what I want to look at is what we as lawyers see in terms of the application of GDPR. When it comes to the application of the GDPR by companies and organizations, there are **some very well-rehearsed issues which people have come across** around **automated decision making** and **obligations to inform people properly,** providing meaningful information. I think those are quite well-known, I will not go into them, I just want to make it clear that these are not the only GDPR issues that apply in relation to AI machine learning. There are lots of others as well.

To give you an idea, focusing on data as an important asset is what we should be doing to enhance our economy. If you look at article 28, these are the rules governing relationships between controllers and processes. This is where one entity is processing data in some way on behalf of another. There are quite **strict rules governing these relationships**. In shorthand, if the entity that is meant to **be just doing the processing** makes some **use of**

the data for its own purposes, rather than for those of its customer, then that **entity becomes a controller**. This effectively puts it in breach of the GDPR and in breach of its contract with the controller. Thus, it inhibits the processor from using this data to train, improve, and optimize the very machine learning systems which are meant to be performing the processing of the data. This will be the case even if, for example, the processor was to fully anonymize the data. Before it did this, it would still potentially be in breach because of the very act of processing data to anonymize it. This is an act of processing which is being done for the benefit of the process. The controller will be back into a breach situation here. I think there are **real tensions between the application of the GDPR and what we are trying to get out of it for society**.

I think it is it is worth just bearing in mind that we should not underestimate the extent to which **GDPR compliance in real life is informed by what you might call market practice**. When there is a **relatively small amount of enforcement activity**, or at least publicized enforcement activity, around certain aspects of GDPR, you **can risk creating a safe attitude** among some organizations. To put it into context, if an organization which might be trying to do the right things sees its competitors behaving in a way which is inconsistent with the GDPR, and then does not see anything being said about this, any regulatory actually being taken, it makes it quite **hard internally for the people who want to do the right thing** to actually **persuade their company to do the right thing**.

Ellis Parry, Data ethics adviser, Information Commissioner's Office

OPENING REMARKS: I am Ellis Parry the Information Commissioner's Data Ethics Adviser. The Information Commissioner is the UK's independent information rights regulator. Elizabeth Denham is committed to increasing the public's trust in what happens to their personal data, this commitment forming the basis of her Information Rights Strategic Plan. The ICO's ambit includes the following legislation, amongst others, the GDPR, the Data Protection Act, Freedom of Information and the Re-use of Public Sector Information Regs. My remit at the ICO is to articulate the interplay between the data protection principles over which the ICO has regulatory oversight and the emerging field of data ethics and communicate and consult on those views; raising awareness and buy-in to the mutually reinforcing nature of the concepts which underpin each discipline.

I have been in post since November 2019. I am a solicitor specialising in information law rights for the past 20 years. Before joining the ICO I was the global data privacy compliance lead for a FTSE 100 pharmaceutical company advising on bioethics and data protection.

MAIN OBJECTIVE: as the rate of data creation increases at an exponentially escalating velocity and the possible use cases involving AI and machine learning evolve too, the undoubted benefits to society of this novel processing and the value it can return is widely recognised. It's been referred to as the "Fourth Industrial Revolution". While the benefits are easy to articulate, so are the risks.

The **EU Commission's White Paper on AI** published last week draws attention to the potential benefits in healthcare, climate change mitigate, predictive maintenance of machines, agriculture, energy, transport and combating environmental degradation. However, as the 2nd recital of the Data Protection Directive opined way back in 1995: "data processing systems are designed to serve man" - technology should serve mankind in order to improve lives while respecting their fundamental rights. There's a degree of alignment about the risks posed by

AI and ML performed on personal data. The following list is not exhaustive but is illustrative of some of the **bigger risks over which there is most consensus**:

- opaque decision making (“black box”)
- entrenching discrimination (data bias)
- unwarranted intrusion into private lives
- challenging human dignity,
- pluralism
- inclusion

“Trustworthy AI” is a prerequisite to its uptake and to unlocking the benefits of this revolution. We engender trust by being **transparent about the risks** and by explaining how those **risks are being mitigated and managed**. If the **risks are identified, mitigated and the success of that governance is subject to ongoing monitoring** the benefits can outweigh the risks.

As the report launched on 10th February by the Committee on Standards in Public Life on “Artificial Standards and Public Standards” makes clear the **Nolan Principles can flex to govern the challenges posed by AI** but AI does pose **challenges in particular Openness (Transparency), Accountability (responsibility for the decisions taken and for the provision of a meaningful explanation) and Objectivity (data bias perpetuating discrimination**).

CSPL recommends the **Gvt issuing an authoritative set of universally applicable ethical principles** while acknowledging all the good work done by the ICO (AI Auditing Framework, Project ExplAIn), the DCMS, CDEI and the OAI/ATI). These **universal ethical principles could bridge any gap between rapidly evolving new technologies and the governance set out for personal data processing in the GDPR** (the text of which was agreed in 2015).

CONCLUDING REMARKS: The rigid construct of black and white letter law may not sit easily with new data processing possibilities leading to a perceived gap in the GDPR’s ability to stimulate and not stifle innovation. **Ethical principles representing flexible consistency through periods of rapid change** can usefully **guide organisations towards a correct operationalisation of the GDPR’s principles even in novel situations**.

Andrew Pakes, Research Director, Prospect

I am the research director of a trade union called Prospect. I am going to talk about work. Prospect is an independent trade union that represents 145.000 members across science specialist professional roles and many workers in technology. My contribution is based on a double insight of our members, because an increasing number of them are involved in the design of machine learning and new technology solutions, particularly those who work in public service and some of the big public service agencies. Secondly, an increasing number of our members are on the receiving end of new technology coming into the workplace. I am going to speak from some of our perspectives in terms of worker experiences and employees.

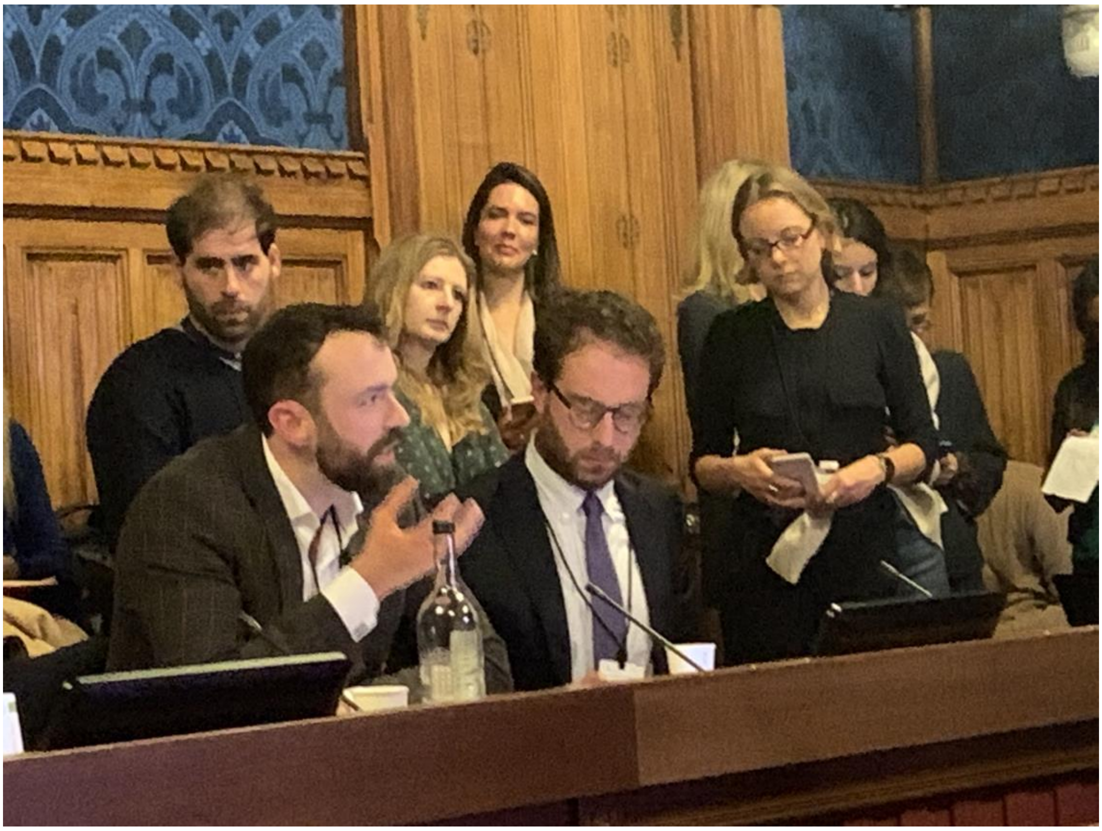
My main message is that in this debate is that **we do not talk about data and work enough**. It seems to be one of the missing areas. When we talk about it, much of that discussion is around data and GDPR. It is primarily based on individual rights and that is important for us as citizens. However, the contractual relationship of work is different to our citizenship rights. It is different to personal data. There is an asymmetry of power, there is aggregate data, there are decisions and consent and choices, which people are asked to make. Individual choices are fundamentally different to me sitting in my house deciding what a costs or b should know about me. The **level of decision-making or power is different to that, so we need to understand how we talk about some of that**. A couple of speakers have talked about value. This is an interesting concept for us. We tend to put value and ownership together, but both of those things lend to have a British end of the debate - which is about individual rights. What we do not seem to talk about much is **value and collective or group rights.** That is important when we look at what this means in terms of work.

The focus that we have been looking at as a union is **people analytics.** How data is changing workspaces. If the 20th century was primarily around regulation between people and management or employers and workers, what we do not have yet is really a **framework which is between people and their data** and how those things go together. The risks if we get this wrong is legend. We have had the tech clash, we have had the stuff around Barclays, we have got the news articles around US companies doing help monitoring and checking how much people sleep at night. All this **datafication and monitoring of work**, unless we have some **concept of group rights**, risks leading to a position where **secondary levels of data, other bits of data, are used to make decisions about people** in terms of their work where there is no balance of power or sharing within that.

Our members are pro technology, we approach this as a union in that we think technology is good. We want to embrace the future of work, but we feel we need to get it right. We did a survey of around seven and a half thousands of our members and specialists in technical roles in January. Half of them said they were **not confident that they knew what data their employers collected about them**, even with GDPR, a third of them were concerned that data was being collected and it may not be used for purposes, and a whopping **75% said that they did not feel they would be involved in any conversations about how technology would change their work** and how that happened. Again, top-down change rather than co-creation. Last week I was at a very similar conversation with a group of businesses and unions in the Nordic countries where they approach this very differently to us. People may be familiar with the analogy of **"accountability responsibility and trust**". Trust is really important concept. I was very taken up by a speaker called Professor Virginia Dignam who asked a series of questions which relate to us and to this work. Accountability, responsibility and trust sound like good things. But **where does consent lie**, **who decides who has a say**, **how do we make choices and trade-offs between conflicting values**, and how do we verify whether design systems embody intended values? **Where do workers sit within that?** We have some very good specialist bodies but if we want to make trust happen, if we want to make change happen at workspaces, there needs to be some **mechanism for workers** - not just unions but workers themselves. There's a range of mechanisms for unions.

I was very taken by the **contrast in the EU approach and the UK approach** and we would be very worried as a union that we will be in a position where we **may end up weakening our commitment to data rights and data governance** in the years ahead, whereas the EU is seeking to build up their rights in trade deals and others. I was taken by this quote in the Digital Strategy last week: "Beyond upskilling, workers and employers are directly affected by the design and use of AI systems in the workplace. The involvement of social partners will be a crucial factor in ensuring a human centered approach to AI at work." If you contrast that to British government policy, to the government's AI strategy, workers are not regarded in official government policy as having any voice in change. We are regarded as people who will receive training as though we are widgets in the economic model rather than we have autonomy and concerns about how our data may be used. My premise would be if we wish to benefit to the greatest degree as a country and as an economy, then we need to look to how they are doing this in Europe, **based on partnership**, than we do compared to America, where they're doing it based on imposition.

James Kingston, Deputy Director, HAT LAB



I should say a couple of words about the exciting venture that is the HAT-LAB. The HAT-LAB is the innovation space and advocacy hub of the Hub of All Things ecosystem. The hub of all things ecosystem is a system designed to promote personal data accounts - the ability of individuals to use personal data accounts to flow all their data into one particular space upon the cloud. This means they can own their data, transact, and use their data from there. Everything we do is designed to promote that vision, both within the HAT-LAB itself as an innovation centre and other spaces within the system. We do that because we see this as the best way of promoting, broadly considered, data rights.

What do we mean when we say data rights? There's **privacy** and we have a number of privacy promoting systems inside the hat. More importantly, however, it means **ownership and mobility**. I'd like to get into that a little bit more. One of the questions that was posed this evening was about **data as a commodity**. This is a potentially problematic idea. **Data is very subjective and immensely private.** It varies incredibly. I think we should consider data as not a commodity, not oil (because you can keep on reusing it), but as a store of value. **Data is a store of value**. What you need for the store of value, this potential source of value I should say, is a **place to store it**.

Where does its value come from? The **value of the data comes from movement** - from your ability to transact and do things with that data. Simply having an inert store of value somewhere where no one can touch it is not going to help you very much. Be you a gigantic level conglomerate or be an individual that is striving to realign the value hierarchy of the internet. To us **ownership means as a classic definition rights and property**. However, more importantly, it means **rights in the property and a sense of agency**. Everything we do in the hat' system is designed to promote this.

Who should own data? What's a model for data ownership? There are a few candidates. At the core you have the **notion of it being either the collector of data or the generator of data** - either the company that is setting the system or the person who's interacting with the company. These are all interesting questions. One of the fascinating things about this system is the way it forces you to consider deep questions of political economy. Like the way people have defined property in the past, we can see it being defined in a similar way today. In the past people have talked about communal data rights, collective data rights, and individual data rights. We now have the concept of the **data commons, data trusts, and individual property rights.**

It is probably clear by now that the HAT is all about promoting individual property rights. We believe this is the best way of shaping an internet that has the interest of individuals and societies at the core. What do we do? **We give people databases** - we give them a space to **access that database right**, and to have that **substantial investment of effort required to earn it**. We give them a **contractual system** to help them along with. We give a **consent driven system for people to consent to potential transactions** to make sure that everything is GDPR compliant and we enable the **co-creation of value between individuals and between the companies of whom they transact**. This is ultimately a common good, just as much, and even better, than any data common could be. We have this notion of whether data can be democratic. I smiled when I saw this question, because ultimately, what is democracy? There are many different forms of democracy. In the HAT **we strongly believe that the best way of ensuring beneficial outcomes for all is by property rights and by spreading property rights widely through society**. This is what will enable the protection of privacy, it will enable individuals to realize value, and ultimately it will enable some sort of redistribution of value.

Professor Ryan Abbott, Professor of Law and Health Sciences, University of Surrey

I would like to thank the All-Party Parliamentary Group for inviting me to present evidence. I have three submissions.

First, it is not clear that ownership is the best way to think about data. It is common for people to speak about owning data and **claims to data ownership are a standard part of many commercial agreements**. But I am not aware of any jurisdiction, including the United Kingdom, in which data per se can be owned. One can have various rights, even **exclusive rights, associated with data**. I may have a right to prevent others from using particular data related to me in certain ways that I find objectionable, a right to prevent others from copying my photograph which consists of data, or a right to prevent others from accessing commercially important data that I keep confidential.

When people speak about **data ownership**, they are often really discussing **particular rights or obligations related to some forms of data**. It conveys a sense of importance in places like here where personal property is a protected interest, but it is **not the right focus in terms of regulation**. It is not the right focus because **ownership involves a level of possession, control, and interest that would be counterproductive.**

My second submission is that data ownership would be problematic because **it is not clear who should own data**, or that the **benefits would outweigh the costs** of any system of data ownership. You could think about the creation of **intellectual property** as usually involving **one layer of complexity**. We have a rule for when a right subsists in something, based on some sort of **activity we want to encourage**. So, copyright can subsist in a novel, and we have a rule for when someone does something that makes them an author. This is because we think that a system of property rights is an important way of encouraging people to write books.

Sometimes that gets more complex, or you have derivate works, or you have machines doing the writing, but **the basic principle is that intellectual property law incentivizes socially valuable behaviors** that would otherwise be inadequately motivated. By contrast, **the development of data tends to be more complicated and involve numerous layers**. Different parties may be generating data, aggregating data, structuring data, transforming data, analyzing data, and using data. It would be challenging to create a workable system of rules to determine which parties in that process would own what data.

It is also not clear what we would be trying to do with those rules that is not already happening. If we want people creating databases or new works of intellectual property, **we already have a system of property rights to encourage this**. If we want people generating data, that is already happening without an obvious need for additional rights. **Data is being created, including by the public, either because it does not require additional effort or because parties receive some benefit from doing so**.

Take the example of data being generated that relates to self-driving cars, which may involve a host of software systems and sensors, inside and outside of a vehicle. Data may be collected by or about a vehicle’s owner, passengers, other vehicles, road conditions, pedestrians, and so forth. That data has value to a variety of stakeholders, such as manufacturers, insurers, and the government. It is not clear here that there is a single party that should own any of that data outside of existing systems of intellectual property rights. It is also not clear that having a party own data will result in more data being collected or data being put to more productive uses.

There are **significant costs to property rights—transaction costs, compliance costs, and administrative costs**. More rights may **increase complexity and barriers to entry**, particularly for SMEs. More rights may also **restrict competition and the rights of third parties.**

My third submission is that **data can deliver public value without being owned**. It can be applied to research that advances knowledge and commercialized in ways that benefit the public. Data can improve a variety of products and services, particularly with respect to artificial intelligence. To the extent **people should be entitled to a greater share of the value being generated from data about them**, and there is a concern that **private industry is disproportionately benefiting from the current system**, the **best means of distributing that value may be taxation**. Ensuring that companies which generate value from UK-based data are paying appropriately into the tax system may re-distribute wealth in a more efficient manner.

To the extent that consumers or the public have concerns about the way their data is collected or used, this can be addressed by **privacy laws**. There is a critical role for regulation with respect to data - less with ownership than with things like **safeguarding privacy, encouraging interoperability, and prohibiting anticompetitive behaviour and discriminatory pricing**.

Dr Tirath Virdee, Ai Capability Lead and Director of Artificial Intelligence, Advisory Board Member of the APPG on AI

Just last week (22nd Feb 2020), Helen Dixon, The Irish Data Protection Officer, stated that her office had hired specialist lawyers to look into punitive financial penalties for technology multinationals as a deterrence for breeching GDPR. The Irish, of course, have been rather slow in their pace of decision making and have faced criticism from some other European privacy regulators. The Irish hold a trump card in so many ways as there are so many technology multinationals with headquarters in the Irish republic. The companies in debate are not just Facebook (including Instagram and WhatsApp), Google, Amazon, but also other nationals and multinationals.

Again, just last week (19th Feb 2020), Ursula van der Leyen, the president of the European Union, unveiled a set of proposals that focuses in on trust and transparency so as to enable the EU to playing field with competitors from U.S. and China. The emphasis is on trust as a mechanism for building a brave new world enabled by AI. The challenges posed by climate change, mobility, and healthcare were emphasised with view to promoting a responsible human-centric approach towards AI. She also said that the proposals should be free from any algorithmic bias. The idea being that whomsoever wants to trade with the EU should respect “our rules and values whilst doing business here”.

EU has been pursuing the idea of a single digital market in the belief that pooling data from governments and businesses would create an asset for Europe that would leverage the possibilities offered by AI. The concern that a handful of corporations control much of this data is touted by the EU as of grave public concern. Margrethe Vestager, the vice chair of the EU for Europe Fit for the Digital Age, says that she wants to ensure that companies of all sizes can compete on equal terms. In addition, the EU is keen to ensure that the rights of the citizen are respected and that profits are taxed where the value is being created. The goals outlined in the whitepaper propose the creation of a framework for trustworthy AI. The EC has also disclosed plans to spend almost $21 billion on AI and data research programs and platforms that should eventually allow for the pooling of data.

Trading in data is crucial for economic growth. The issue is political, economic, ethical and even moral (in that it very much impinges on our personal privacy and security). It is the new oil and it powers the digital economy. Smart (IoT) devices generate and collect a wealth of personal and enterprise data. These in turn pose serious legal and ethical issues especially if this data needs to be traded. The issue to **consider firstly is whether we are talking about enterprise data or personal data.** As Václav Janeček, of Faculty of Law at University of Oxford points out, we need to **distinguish between 'data' and 'information'** – he points out that there are constructive limits of personal data ownership; especially ownership of data from IoT.

There are two models of data ownership for enterprise data:

- **Assigned data ownership** - This approach acknowledges that enterprise data is "owned" by the enterprise rather than individuals or groups within the enterprise. Accountabilities for working with that data is assigned to roles in the organization, and individuals or groups fill these roles. So, it is often convenient to designate "Data Owners" who coordinate accountability.

- **Federated responsibilities -** enterprise dataflows through an organization, touching many business and technical processes and being stored/moved/transformed by many IT systems. Federated responsibility dictates that data lineage (the path data has taken from its creation/acquisition to a specific system or report) first be documented and then assign data-related accountabilities for a manageable number of segments to Data Stewards, SMEs, and/or Data Custodians (technical resources).

Taking back control may have been the battle cry for Brexit, but personal data poses a number of problems related to ownership. Individuals are now starting the drum beat of 'take control back'. We are faced daily with terms and conditions and the automaton in us wants what we want, and we cannot be bothered with reading the often-onerous conditions for participation. We are not usually directed to a remuneration page for us trading our data and we realise that sometimes the free nature of their services is at the expense of our privacy. **Only extensive legislation can result in the property rights related to personal data;** the fact that one lends one's personal 'art work' should not let the borrower claim permanent rights, nay ownership, of it. The legal innovation required around this is the next logical step after the General Data Protection Regulation (GDPR) which came into force in May 2018 in the EU. Significant economic and political battles are ahead on this most crucial of arenas. Companies such as Capita may build technologies that enable post-GDPR compliance and personal-data de-identification in our cloud and data-lakes, but not all companies are that responsible in a corporate sense.

References:

https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf

3. Beyond GDPR - Recommendations for policymakers

The evidence presented by our expert speakers highlighted important issues in data governance, especially in the implementation and enforcement of the GDPR, that obstruct the responsible collection, processing, and sharing of personal data.

appg
All Party Parliamentary Group on Artificial Intelligence
EXPLORING THE IMPACT AND IMPLICATIONS OF ARTIFICIAL INTELLIGENCE

Consent:

Under the GDPR the data subject has to give **meaningful and informed consent** to the processing of their data. It is however questionable if the current procedure of presenting **several pages of consent forms** is helpful in this regard as they are unlikely to be entirely read and understood. Meaningful consent must be **based on precise communication** that is **easily understandable**. Evidence presented on this topic suggested the design of **best practices for an effective communication** to obtain informed consent.

Beneficiaries of data sharing:

It has been suggested at the meeting that personal data governance debates tend to be too focused on the issues of control and consent whereas **questions for the beneficiaries of data sharing are often overlooked**. Consumers must be informed who will benefit, in **economic or non-economic terms**, from their data. Further, how can consumers be ensured that their data is only being used for **benevolent purposes**? This is especially critical in the case of **highly personal data** whose sharing might have **detrimental consequences** to the individual.

Data privacy:

Several models for enhancing data privacy, ranging from **data commons to data trusts to personal data accounts**, have been introduced at the meeting. At the heart of each of these methods lies the idea of **responsible data sharing** by granting individuals **agency over the sharing of personal data.** However, these methods and their benefits must be **communicated to consumers effectively** to gain traction. Further, there must be more **clarity about the risks and implications** of the deployment of privacy enhancing tech.

Hidden data:

Currently there still exists **significant barriers to reaching potentially valuable data**. These are usually generated as an **unnoticed side-effect** in business operations of companies that have not yet realised the value of the data they collect. These companies **require assistance in retrieving the valuable data** they accumulate to improve business operations and services.

Enforcement of the GDPR:

Lastly, it has been criticised at the meeting that so far the **enforcement of the GDPR has been rather patchy**. As a consequence, companies which are abiding to the GDPR will face **problems staying competitive**. That is, GDPR compliance in real life, as was stated by one of the expert speakers, is currently informed by **market practice**. Further, it appears **GDPR compliance towards employees** is **not fully reallised** in many work environments. The processes and objectives of **data collection and monitoring** at the workplace are often not satisfactorily communicated to employees. This can create **privacy concerns and fears of surveillance** at the work place. **Efficient communication** and an **open discussion** of the implementation of data collecting and processing technologies at the workplace are thus indispensable.

Contact

APPG AI Secretariat

Big Innovation Centre
62 Wilson Street
London EC2A 2BU
United Kingdom

info@biginnovationcentre.com
www.biginnovationcentre.com

appg@biginnovationcentre.com
www.appg-ai.org

www.biginnovationcentre.com