High Level Policy Hearing:
'Building a European Data Economy'

Transcript

Tuesday, March 28 2017
Held at: Salle 11, 11th Floor Berlaymont, 14h00 to 16h00
European Commission, Rue de la Loi 200

Participants:

MODERATORS:

ANN METTLER, Head, European Political Strategy Centre

And

MARIO MARINIELLO, Digital Adviser to the European Political Strategy Centre

PANELISTS:

Ciro CATTUTO, Scientific Director, Institute for Scientific Interchange Foundation

Aija LEIPONEN, Associate Professor, Applied Economics and Management, Cornell University

Gerald SPINDLER, Professor for Civil Law, Commercial and Economic Law, Comparative Law, Multimedia and Telecommunication Law, University of Goettingen

Maurice E. STUCKE, Professor of Law, University of Tennessee [by videoconference]

Peter SWIRE, Huang Professor of Law and Ethics, College of Business Georgia Institute of Technology [by videoconference]

Christiane WENDEHORST, Professor of Law, University of Vienna
Good afternoon and welcome to this hearing convened by the European Political Strategy Centre on 'Building a European Data Economy'. My sincerest apologies that we are starting late, we were trying to connect with one of the external experts, Professor Swire from the Georgia Institute of Technology. Hopefully we will still manage to do so, but this explains the delay. My name, particularly for the colleagues joining us from abroad, is Ann Mettler. I am the head of the European Political Strategy Centre, the European Commission’s in-house think tank. I am joined on my left by Mario Mariniello, Digital Advisor to the EPSC. This hearing is organised to contribute to the European Commission’s consultation on the data economy and everything that is being said here today will be transcribed and submitted to this consultation.

Let me first of all say how delighted I am to welcome such a high level group of experts who will help us today shed light on the data economy. They are in alphabetical order, Ciro Cattuto, Head of Data Science Laboratory [Scientific Director] ISI Foundation. Aija Leiponen, Associate Professor of Applied Economics and Management at Cornell University. Gerald Spindler, Professor of Civil Law, Commercial and Economic Law, Comparative Law, Multimedia-and-Telecommunication Law at the University of Goettingen in Germany. Professor Maurice Stucke, Professor of law at the University of Tennessee who is joining us via video link. Hopefully with us soon is Peter Swire, Huang Professor of Law and Ethics at the College of Business, Georgia Institute of Technology. And lastly, Professor Christiane Wendehorst, Professor of Law at the University of Vienna. As you can see, four of our experts are in the room, and one and hopefully soon two, of the others are joining us from the US.

Now before I ask the invited guests to briefly introduce themselves and their work, allow me to make a few announcements. The hearing will last about two hours and each speaker will have a certain amount of time to address each question. The time limit will be announced when the question is posed and one minute prior to your time being up, we will show you an orange sign and when your time is up we will show you a red sign, so at that time I will really then have ask you to wrap up. Some of the experts have said well, I mean, I need a little bit more time for one question and I think in principle that would be fine as long as you then subtract some time for some of the other questions that you will answer so we can more or less stay on track with the time table.

As I said before, the hearing will be on the record and a full transcript of the hearing will be submitted as a contribution to the public consultation on data. The experts have received the questions in advance in order to allow them to prepare well their answers. We have invited a few colleagues from the European Commission to join us as observers and they are the people sitting behind me and the people sitting to the side of me. However given the format I would appreciate if our Commission guests can be in full listening mode for the duration of the hearing. There will be an opportunity to interact with the guests, later at the networking coffee that we are organising, after the hearing is over.
And with that, let’s go to the first question for which you have a maximum of one minute please, one minute. So the question is:

‘Please state your name and affiliation, please flag any potential conflict of interest you might have and please describe your background and your experience in dealing with the Data Economy from a public policy perspective’.

We first go to Doctor Cattuto, one minute or less please.

**CIRO CATTUTO**

Good afternoon, thanks for having me here, my name is Ciro Cattuto I am the Scientific Director of ISI foundation, which is a non-profit privately funded research institution in Torino, Italy and New York City, USA. My background is in physics, I have a PhD in Physics, I founded and lead Data Science Laboratory at my institution. Most of my work focuses on using digital data, in particular proxies of human behaviours to inform models, ideally predictive models, for health and infectious disease dynamics. I work with sensors, I work in data science, and more recently I have been investigating with my research team the impact of data science for public good and international development.

**ANN METTLER**

Very good, thank you so much. Second is Aija Leiponen. Professor Leiponen, one minute or less please.

**AIJA LEIPONEN**

Thank you, I'm Aija Leiponen from the Dyson School of Applied Economics and Management at Cornell University. I study and teach Innovation and Technology Strategy. I have been teaching digital business strategy for fifteen years at Cornell. My interest in innovation has been particularly in digital industries, and have studied communication technologies in particular. For the last three or four years, I have studied the underpinnings of data markets and innovation in the data economy with colleagues at Imperial College London. So a lot of this work has been centred in the UK.

**ANN METTLER**

Excellent, thank you so much. Third is Gerald Spindler, Professor Spindler please.
GERALD SPINDLER

My name is Gerald Spindler, from the University of Goettingen, Germany. I'm a lawyer as well as an economist. I'm actually a Professor in the faculty of Law in Goettingen. I have been doing, as you have heard, comparative law as well as corporate law as well as everything which is related to Internet; in particular liability issues and intellectual property, for example, as well. There is no conflict of interest, as far as I've seen. My background is that I, have been, more or less in the Internet industry since 1996, the very beginning of the first problems of liability of Internet intermediaries. We've done the 2006/2007 review of the e-commerce directive for the European Commission. We are actually carrying through different European projects, funded by the European Commission, such as data protection and intellectual property rights, and I have been consulting also the German government which I hope is not a conflict of interest.

[Laughter]

ANN METTLER

Thank you so much. Then fourth is Maurice Stucke, I hope I pronounced that correctly. Professor Stucke, over to you please.

MURICE E. STUCKE

Yes, my name is Maurice Stucke I'm a Professor at the University of Tennesse. I'm also co-founder of the Konkurrenz Group, a law firm. I worked for many years at the US Department of Justice, in their anti-trust division. More recently I have been the co-author of two books on Big Data. First is 'Big Data and Competition Policy' and the second is 'Virtual Competition'. There are no conflicts.

ANN METTLER

Excellent, thank you so much. Then last, but not least, is Christiane Wendehorst. Professor Wendehorst, over to you.

CHRISTIANE WENDEHORST

Thank you very much. My name is Christiane Wendehorst, I am Professor of Private Law at the University of Vienna, and currently also the Vice-President of the European Law Institute (ELI). I teach private law, and my research focus has recently been on the challenges posed by digitalisation, *inter alia* the Internet of Things, artificial intelligence and the data economy. Like Professor Spindler I have also given advice to the German government, which I do not consider to be a conflict of interest. I cannot see any other conflicts of interest. I am currently leading a joint
project between the ELI and the American Law Institute, together with colleagues from the US on how to adapt laws to work in the data economy. Thank you very much.

MARIO MARINIELLO

So, let me now read question two. This is about the context and each speaker will now have three minutes to reply. And the question is:

“What are your general views on global trends linked to the emergence of the Big Data paradigm? And what is your assessment of the European Union’s progress towards a data economy?”

And this time we start with Professor Leiponen.

AIJA LEIPONEN

Regulation issues in non-personal data, primarily, I view it as an innovation problem. I think Europe has a lot of challenges in creating services and technologies around data that are currently not there. The policy challenge is to create a governance system that would provide optimal incentives for innovation and creating that data economy. We know that the technological frontier is moving very fast, particularly in software-based analytics, including artificial intelligence techniques. And we know a little bit of where that may be headed, but it’s evolving very rapidly right now.

In the future, or the near future, much more of the data that is currently being collected will actually be analysed and utilised for decision making. Currently, overall, globally, we collect a lot of data and store a lot of data that is actually never touched afterwards. So once we develop and adopt a lot more of the analytical tools, more data will be utilised for decision-making. And when that happens, more routine decision-making will be automated, and that might include decisions in such areas as recruitment, investment, or administration. So a lot of activities where there is repetition can be automated, whereas human judgement will be critical in other areas such as where creativity or emotional intelligence, caring for other humans, will be important. So that’s going to be the human specialty.

We should be thinking about this not necessarily as a data economy but as an intelligence economy, where data are strongly complementary with analytical techniques, artificial intelligence and software accompanying those tools. Basically what we need is training datasets for intelligence, and we need intelligence in applications and algorithms to make sense of the data. Those things have to go together.

At this point, I see few areas where there are likely to be data monopolies, but I think control and market power in that intelligence economy will be in platforms and in the
intelligence technologies. And data will be the fuel that keeps the system running. And where I see European weakness, is particularly in the adoption and application of computer science, advanced techniques in computer science, in commercial applications. Not necessarily of the leading edge science, but the application of those ideas in commercial settings.

MARIO MARINIELLO

Okay, thank you very much. I’m just told by Ann that Professor Swire can hear us, but we cannot see him...

PETER SWIRE

Yeah, sorry, the video, I wasn’t able to make it work. I apologize.

MARIO MARINIELLO

Yeah, sorry, sorry also from our side. I suggest we will continue to answer to the second question, and then you will be able to reply to the first question which is essentially introducing yourself altogether. So we will cover from there. We now go to our next speaker, who is Professor Spindler.

GERALD SPINDLER

Thank you. There is something to add, from a more legal point to what my colleague Leiponen already stated, from a more economic point of view. It surely is the interplay with artificial intelligence, which will play as a technology an eminent role in the coming years. As we’ve seen already in the other discussions, there is a close connection between Big Data, data and upcoming algorithms, artificial intelligence, etc. So the impact of data analytics on our society is very clear, and you can see it in every sector - be it logistics, be it insurance, be it healthcare, etc. And even in the so-called, well this is a typical German notion, industry 4.0, which means a very strong collaboration, cooperation between however independent partners in closed networks. These networks are based on a common platform or sharing of data. So, the impact of data analytics and algorithms is very clear here.

So what is the role of public policy in the data economy? As an economist, I have to state that I strongly believe that the state should only intervene when there is a market failure. That is to say, markets are best suited to adopt a fine-tuned solution, and only then, when that doesn’t sort out, we have to step in. So we really have to do a lot of empirical work on that in order to
see whether there is a market failure or not. But then surely, we have to intervene, be it in anti-trust law or whatever.

What are European strengths and weaknesses etc.? I think, contrary to what industry is always saying, data protection could really be some of the advantages that we have in creating trust for the people using Internet, services, etc.

One of the weaknesses for sure is, and this is just an example, of the European Union the scattered landscape concerning the regulations being in place at the national level and concerning in particular the copyright law, this is absolutely the inflexibility we have of copyright law.

MARIO MARINIELLO

Thank you. Professor Stucke, over to you.

MAURICE E. STUCKE

Yes, we are looking at these issues from a competition policy perspective, and we are looking both back and looking forwards. Looking back, we’re increasingly realising in the US, the failures of US competition policy over the past thirty years. In April 2016, the White House issued an executive order and report on the state of competition in the United States, and the report identified several disturbing trends, since the 1970s. Competition appears to be decreasing in many US economic sectors, including a decade’s long decline in the number of businesses being started and the rate at which workers are changing jobs. At the same time, many industries are becoming more concentrated, with profits increasingly falling in the hands of fewer firms. So the solution is more competition, which judicially means more robust anti-trust enforcement.

Looking forward, what does this mean with the data-driven economy, and what does this mean for building a European data economy? My short answer is this: We want to ensure that we have the policies in place to maximise the benefits of a data-driven economy, while mitigating the risks. So the goal should not be simply to maximise the number of cloud service providers in the US or the number of super platforms to compete against GAFA: Google, Amazon, Facebook or Apple. Rather the goal should be to promote a data-driven economy that is inclusive, that protects the privacy interests of the citizens, promotes citizens’ well-being and promotes a healthy economy. Here Big Data and big analytics can promote competition and our welfare by making more information, more easily available and providing easier access to markets. But we can’t assume uncritically that we will always benefit.

At times, Big Data and big analytics can be used to defy competition. As we discuss in our books, Big Data and big analytics can lead to anti-competitive outcomes such as innovative ways for companies to collude, innovative ways for companies to price-discriminate, innovative ways for dominant firms to abuse their position, and anti-competitive data-driven mergers. I will touch on these topics in response to the subsequent questions. Thank you.
MARIO MARINIELLO

Great, thank you very much. Professor Swire, can you hear me?

PETER SWIRE

Yes I can. Can you hear me?

MARIO MARINIELLO

Yeah, we can loud and clear. If you could be so kind as to also address the first question now, introducing yourself first and then address the second question that I just read?

PETER SWIRE

Yes, thank you very much. And thank you for the opportunity to participate in this. In terms of my background, today I'm the Huang Professor of Law and Ethics at Georgia Tech Sheller College of Business. I also have appointments in the College of Computing and in public policy. I have been working on privacy and Internet issues for more than twenty years. In 1998 I wrote a book on EU-US data protection and that led up to my participation in the Safe Harbour negotiations in 2000 when I was President Clinton's chief counsel for privacy. During that time I also worked on many other issues, I was the White House coordinator for the Health Insurance Portability and Accountability Act (HIPAA) medical privacy rule and went through 50,000 public comments and came up with a rule that’s been, I think, pretty stable and successful.

I have continued to work on many of these topics for many years. In 2013, after the Snowden revelations came out, President Obama named me one of five people for the Review Group on Intelligence and Communications Technology, sometimes called the National Security Agency Review Group. And so in that realm I got to see a lot about the intelligence community collection of Big Data and related things. I also, and we will get to this later in the discussion, have written a long law review article about Data Portability under the EU proposed, at that point, proposed general protection regulation, but we’ll get to data portability later in the discussion. So, I come to this with a lot of experience in privacy and cyber-security in European data protection, but also I’ve been trying to understand Big Data in various parts of my research.

So, that is background, the question is what are the strengths and weaknesses of the European Union in connection with Big Data? And I think that there is a challenge in the economy. The data protection regime in Europe sometimes is seen as having the protective principle, which is the idea that there is a presumption against processing of data under General Data Protection Regulation...
Protection Regulation and other directives. The presumption is to protect fundamental human rights. Many times I agree, there should be more care and attention to these protections than US law gives. But I think when it comes to Big Data, it means there’s a sort of initial presumption against using data. In the United States the presumption is that it’s okay to use Big Data, and in China even more so, I think, if we look at where new possibilities of work come from.

And so, that initial leaning towards protection rather than initial leaning towards processing data plays into the question of who gets to be first mover for innovation, and in many parts of the information economy for reasons that, I think, that reports of the EC show, first mover is often who can achieve scale, often get an advantage. Then there’s network effects, there’s tipping, cascades and all the rest. And so, for information processing industries, I think it’s been true for the last set of years, 10-15 years, that few of those that have ended up on a global scale succeeding have come from Europe. And first movers have tended to come from the United States, and in some instances China. That’s not a comfortable conclusion for Europe, because fundamental rights and data protection have become such important projects to what the European Union means. I think if you’re going to have a realistic discussion about Europe and Big Data, if Europe is rarely the first mover, it’s going to be a difficult challenge.

Now, one of the big shifts is that a lot of the activity, again as shown in the EC reports, in the next period of time is going to be about Big Data where personal data is either not there or is not the leading thing, when you think about industrial robots, or machine tools, or some of the fantastic tooling done in Germany and many other countries. And so how to take Europe’s traditional strength in these areas and succeed in an era where sensors have become pervasive and cheap? That’s a challenge of how to not have the presumption against data, and I think it’s an uncomfortable discussion, because we’d like there to be trust, and we’d like there to be fundamental rights protection. But, I think realistically, looking at the economic effects, Europe will be missing something important if it thought that trailing, not being first mover, wasn’t a big problem. I’ll stop there, Thank you.

MARIO MARINIELLO

Thank you very much. Exactly on time [chuckle]. Professor Wendehorst.

CHRISTIANE WENDEHORST

Thank you very much. I think data-driven innovation is expected to have a huge impact on almost all aspects of society and the economy. And it has already been mentioned that it’s not just the data economy. The Internet of Things, Artificial Intelligence, all these link together, and data is probably key to everything. It may also mean that we will have to rethink much of what we have taken for granted for quite a long time, and reconsider the way the law and the economy works.
Europe, as Professor Swire has just pointed out, has started bracing itself for the new era with the General Data Protection Regulation, and data protection has an international reputation for being extremely strong in the EU. Now don’t get me wrong, I am a consumer lawyer, and I believe data protection is an extremely important value. And when I am discussing with my US colleagues I frequently get the message: ‘Well sometimes we wish we had something like European data protection.’

However, even if it is a strong and important signal internationally, it is of course not precisely a signal pointing towards Europe as the world hub for the data economy. Just look at the European Data Protection Supervisor statement of two weeks ago, which was indirectly comparing any economic transaction with regard to personal data to trade in human organs. [chuckle]. This is the kind of attitude that is transported internationally. So I believe that a major challenge Europe will have to face, is to reconcile personal data protection - which must remain extremely strong, don’t get me wrong on that - with the goals of the data economy.

I am firmly convinced the two need not contradict each other, but it will not be easy to get the balance right. The data economy is a particularly difficult area where to regulate. This has several reasons; it is extremely innovative, fast-moving, and it is opaque because data are intangible, invisible, not registered, and often secret. So I think a lot of challenges are lying ahead. Personally I believe that we may be well advised not to rush things, rather to wait how things develop, maybe to make some changes in contract law, to focus on competition law, including on strong enforcement - small changes that may have big effects, but not to rush things in order not to impede the European data economy. Thank you very much.

MARIO MARINIELLO

Thank you very much. Dr. Cattuto.

CIRO CATTUTO

Thank you. I’d like to share a few comments in general on global trends related to the adoption of data technologies. What we are going to see going forward, I think, is a series of unprecedented technical capabilities, starting from a capability to quantify and measure reality. I would say that, if we should sum it up, the revolution driving the data economy will be that the digital image of the world, the digital image of reality, will track reality closer and closer in terms of granularity, resolution of individual behaviours, timeliness.

So, this is one trend. The second trend is a push towards decentralisation and distribution. The Internet of Things will drive this forward very fast. Simultaneously, the adoption of machine learning technologies and Artificial Intelligence. Artificial Intelligence meant in the soft
form, in the form of advanced machine perception - not general AI. This will create a series of intelligent, semi-intelligent, or smart agents, powering products and services, and these agents will be endowed with part of our agency as citizens, institutions and consumers, and they will engage in transactions on our behalf.

So we are looking at measurability, quantifiability, decentralisation, autonomy. All of this will converge on creating a landscape where we will have more and more algorithmic capability and algorithmic decision making in the loop of society, in the loop of the systems of society. And these kinds of capabilities will span the public sector and the commercial sector, because they are fuelled by the same data, and because they tackle the same kind of challenges.

Another thing that will happen looking forward is the vanishing of the perception of interaction with these technologies. All of the touch points with technologies will vanish because we will have more and more ambient technologies and more and more ambient intelligence due to the Internet of Things, which means that many boundaries between the public sphere and the private sphere will be blurred in their definition. We’ll be giving commands to our personal devices in public spaces that can recognise us as citizens, as individuals.

So, in all this, the impact of analytics is transformational, because, and I agree with the colleague about this, their value lies in intelligence. The analytics are really what allows us to transition from data to a signal, a high-level signal upon which we can take a decision. And eventually the data economy is about an ecosystem of high level signals that are actionable in our capability to use them to take decisions about our market systems.

Now I believe, and this is my key point here, that market forces alone will not deliver the full impact of these technologies. Because the impact needs to be aligned with the mission, the priorities and the values of the people and the institutions taking the decisions. So it’s very close to the top of the organisation using the data, so market forces will deliver products, will deliver enablers, but the way we use data through analytics to take decisions, that eventually will be a missed opportunity for the public sector unless we use legislation, regulation and a variety of other means, moral suasion, corporate responsibility, all the means we have to encourage institutions and organisations to allow this market to benefit the public sphere. So, to sum it up, I believe that public policy strongly needs to encourage, facilitate, and incentivise the creation of shared data assets as well as the public impact of data on research and on public interest in general.
Based on your professional experience and research, do you believe that the European Commission is right to deploy a set of initiatives beyond what has already been done, such as for instance the General Data Protection Regulation, beyond that aimed at improving policy and legal framework for the data economy, in particular as regards access to data for use, reuse and transfer and ensuring the free flow of data inside the EU? Please provide your assessment of the European Commissions' proposed initiative and indicate which areas of intervention should be prioritised and why?

So again, five minutes. And we start with Professor Spindler please.

GERALD SPINDLER

Thank you. I think your questions and the more specific questions are divided in three sections. The first is aiming at the data protection regulation. The second, at data localisation and, third, access to data, which will then be more discussed in part 4 of the core questions.

So firstly, concerning the General Data Protection Regulation, I think there is a huge amount of non-personal data there in the world, depending largely on the definition of what personal data is. And there the lines are still blurred and we still need some more guidance on that. To be realistic, it can only be done by the institutions which have already been established by the General Data Protection Regulation, in particular the European Data Protection Board, today the Article 29 Working-group. But there could be substantial support by the European Commission, given for example, guidance or research projects in order to develop standards when we talk about anonymised data, which is out of the range of the General Data Protection Regulation, and this is absolutely necessary.

The second part, where the European Data Protection Board could be a substantial help as well as the Commission is to define the tricky question of consent – of consent to data processing – which is quite opaque in the General Data Protection Regulation, well of course due to political discussions, and especially concerning the so-called tying clauses. So there is much that has to be specified under the General Data Protection Regulation and concerns some sub-legal innovations that you can do without rising or stirring up too much discussion on the political element.

So, secondly, concerning the data localisation, I can just state that of course under the European Treaty, there is always the need to justify restrictions on the free flow of data such as issues of sovereignty, of national security which may start here, which may be qualified as a justified restriction. However, everything else, even in tax law etc. I would be extremely doubtful whether this could be a justification for data localisation.

Third point, access to data, I would really like to move this to my part of the answer for the core question number 4 because I think it is strongly related to the other question which goes more into detail, so I spare a bit of my time for that.
Indeed, we’ll add it then to question number four. So, next up is Professor Stucke please.

MAURICE E. STUCKE

Yes I agree with the initiatives, with a couple of caveats that I will touch upon in response to question four.

As a general matter, data is a key input in the data-driven economy. A lot of attention has been paid on personal data, but non-personal data can be critical as well. Companies are increasingly undertaking data-driven strategies to obtain and sustain a competitive advantage. As we discuss in our book ‘Big Data and Competition Policy’, firms are already securing significant returns from their Big Data investment. So you need to consider holistically how you can promote this data-driven economy and the ability of EU firms to compete in this economy. The European Commission Staff Working Document, dated January 2017, identifies several important mechanisms. One key avenue to improve the free flow of data is to improve the current legal institutions. This would involve clarifying issues of ownership rights of non-personal data. It would also include streamlining the ease with which parties can transfer data via contracts.

The second key avenue is removing welfare reducing governmental restraints on the free flow of non-personal data. One concern you identify is data localisation. You want to ensure that any current or new data location restrictions are justified. I would encourage you to ask these three questions. First, whether the national governments’ expressed interest is substantial. Second, whether the state action directly promotes that substantial interest, and third, whether the state action is more extensive than necessary to promote that substantial interest.

Even if you improve the legal institutions, and even if you remove the unnecessary public governmental constraints on the free flow of data, I agree with Ciro, that you cannot necessarily assume that market forces will efficiently allocate the non-personal data. So one area of intervention that deserves more attention is the role of competition policy in promoting the free flow of data, and how market power can impede the free flow of data. Normally when we think of market power we think of prices, namely a firm’s ability to raise price above the competitive level. But in a data-driven economy, firms can exercise market power by collecting more data than they otherwise could at a lower price than what they would otherwise pay, and they could also restrict others from accessing this data. One example would be farmers, and a few powerful farm equipment manufacturers. Farmers create the raw data, but the data automatically goes to the manufacturer: since it’s non-personal data, the General Data Protection Regulation does not apply. Nonetheless the data remains in the manufacturer’s silo, and this can adversely affect
public welfare. So you need to consider then, what are the factors that can lead to market power. One factor may be anti-competitive data-driven mergers; second are abuses by dominant firms. Dominant firms have the data and use exclusionary means to prevent others from accessing the data. Third, are vertical private restraints, for example, manufacturers limit the extent to which others in the supply chain can distribute the non-personal data. For example, the farmers here can only provide their non-personal data to the tractor manufacturers and no one else. Fourth would be anti-competitive actions by key gate-keepers that affect sellers upstream. And two areas that I would encourage you to examine are e-monopsony and e-scraping; we are currently looking at these issues as well. Fifth and finally would be how market forces themselves can limit the free flow of data. This would involve examining at least four data-driven network effects which I will expand upon in response to question number four. Thank you.

ANN METTLER

Thanks a lot. Thank you so much. Next up is Professor Swire, please.

PETER SWIRE

Yes, thank you. I have five points which I will make briefly.

The first is that the emphasis of the single market for flows of data seems to me a very good idea, clearly an emphasis of this entire effort, because many of the barriers to flows of information turn out to be, if not pre-textual, not convincing on closer examination.

A second point in the list of questions has to do with the important categories of non-personal data. And that’s clearly the case, but the non-personal and personal increasingly get mixed. So think about the car industry, where Europe has very strong car manufacturers. Historically that wasn’t an area that involved much personal data about the individual car except who bought it, but the connected car is going forward and we are doing projects in my class this semester about this. A lot of the innovation, a lot of the leadership in the auto industry going forward, is going to be data related, both safety related for vehicle to vehicle information but also it’s going to go to the person’s individual activities whether it comes to music or where they drive or whatever. And so that means that the non-personal is going to get mixed with the personal much more pervasively and that sectors that never thought of themselves in the privacy area now have a lot bigger concern.

The third point then comes to the topic of de-identification, so once there is machine data and other data about cars for example. Then we do have some risk if there is public release of the data, that people can use each of those data points as clues that might in some cases re-identify people. So one thing in my own work, I have emphasised is, public release often does have privacy risks but instead you can create organisational structures where
people have contractual permission. So for instance, in New York City, they have a Big Data
initiative for the city where the agency that does the analytics contracts with each of the data
sources, keeps them confidential, and doesn’t keep the data after that. And that allows a Big Data
initiative on a one-off basis for each project, but it allows you to respect the medical rules or other
rules for each databases. And those kinds of organisational controls I think will be
increasingly important in order to merge data, because if you just put it up on the web, this re-
identification problem is so pervasive.

The fourth point has to do with data localisation and law-enforcement access. This is an
area where a major research project, we’re having a conference here at Georgia Tech on April 18th,
and Bruno Gencarelli, who is Head of Unit for data protection will be the key-note speaker, and with
DG Justice. There are very hard data problems facing police because of encryption and data at rest,
in the cloud in foreign countries, and data in transit, because they cannot do wire-taps increasingly.
And so, police are feeling much more pressure to hold data locally. It’s an enormous
pressure, I think growing over time. So, reform and mutual legal assistance seems to be a much
bigger part of the problem for data localisation than many have recognised, and we’re talking about
ways to fix that.

And fifth I do have a proposal for institutional change, or institutional thought for the
European Union and its various committees. It has to do with finding ways for data
protection experts, economists, Big Data experts, and others to engage each other in a
more systematic way. When I worked in the White House, they had what is called the clearance
process and so we had the privacy people in the room with the Department of Justice, and – in the
room if we needed to – with the National Security Agency. My experience in Europe, including
during the negotiation of the privacy shield, was that there is not the same kind of systematic
and intense engagement between the experts in these different pieces. They are seen as
different projects, and that means that there is sometimes a lack of understanding or a lack of
ability to weigh off the reasonable requests of different perspectives. And so a better mechanism
across subject matter expertise, I think, might be important and in the absence of that, the
independence of the data protection commissioners doesn’t give them much reason to find ways to
get to outcomes that both protect privacy and achieve other goals. So institutional reform to get
the perspectives together, I think, is a much bigger issue than I have heard discussed
previously in the European discussion.

Thank you so much. Then we go to Professor Wendehorst please.

Thank you very much. Looking at the European Commission’s data plan and priorities as they
emerge from the documents dated 10th of January, a central differentiation is made between personal data on the one hand and non-personal data on the other. While I believe this is really a central differentiation one has to make, I also see problems. And some of the problems relate to the facts that have already been mentioned, i.e. that the divide between personal and non-personal becomes blurred. The definition we have under the General Data Protection Regulation is extremely broad, so that means that data that maybe considered anonymous today can well be considered as personal data tomorrow. So the line between the two is a moving target.

That is also a problem with data relating primarily to a business. We are always speaking about machine-generated data, and machines are often used in businesses. However, also business related data may under certain circumstances be personal data, e.g. if they relate to a particular person running a small business - I know this is disputed but it not entirely clear to what extent we are seeing some General Data Protection Regulation restrictions here.

Then there is Member State legislation that includes legal entities in the scheme of data protection. This is not the case under the General Data Protection Regulation, but it is the case at national level. And I agree with Professor Swire that personal and non-personal data gets mixed. When you look at car data, these are arguably all personal data. The reason why there is a good reason for working with personal data is that anonymisation reduces the analytical value of data. So non-personal data do exist but the line is difficult to draw and the line may be moving. This means uncertainty for businesses.

What are the conclusions to draw from all this? Well, first of all, I agree with Professor Spindler that we need more guidance, and that it is probably something for the Article 29 Group or for whomever to provide more guidance as to what counts as anonymisation. But then, speaking about the data economy I think it is wrong to say from the outset we can never include personal data in anything like the data economy. I think that would be the wrong signal. The US and China and other countries do not have these restrictions, and I do not think Europe should say that from the outset. So, the idea should rather be to find ways how to reconcile strong data protection and the goal of having a vibrant European data economy.

Some of the possible approaches have already been mentioned by Professor Swire. I would like to add one further suggestion by myself, which I am aware may be controversial and which I call ‘data trusteeship’. The idea is that we support the development of sophisticated Personal Information Management Systems, also called PIMS, and that those sophisticated PIMS may receive a mandate from data subjects to exercise those data subjects’ rights under the General Data Protection Regulation, plus rights under copyright law. This mandate would be partly non-exclusive, partly exclusive and the data trustees, as I call them, would be in a position to, make transactions with third parties on behalf of the data subjects, but according to standardised directions given by the data subjects - like "no profiling", "only for this and that purpose", etc. – and in the interest of the data subjects.
I believe this could, if designed well, create a win-win situation. Why would it be a win situation for the data subjects? Well, the data subjects would have a single point of contact and they would have an entity that really has the technical knowledge to assess whether purposes-limitations are kept. May I run one minute over? [chuckle] Data trustees would have the technical knowledge to really assess the way data are used and they would be in a position to build up bargaining power and to operate in the interest of data subjects. On the other hand, I think this would be a win situation for the data economy, because it would mean personal data are not from the outset excluded from the data economy, rather it can be part of the new data economy in a way that respects data protection and respects fundamental rights of data subjects. Looking at the time, I will not say anything on data localisation. Thank you very much.

ANN METTLER

Excellent, thank you so much. Next is Doctor Cattuto, please.

CIRO CATTUTO

I would like to comment on the fact that in the data economy there is a huge untapped potential of usage of non-personal data for decision making, mostly, right now, due to silos. Non-personal data allows the creation of valuable data assets that can be assimilated to maps. If you think of topographic maps, they represent knowledge about space, and this knowledge is shared and enables decision making about space, and movement in space. In the data economy, Big Data sources will afford exactly the same. We will use mobility maps from mobile phone data, energy consumption, data, etc., to map out poverty or welfare. We will use financial transactions at points of sale to draw the map of a hidden geography of who buys what products at which location.

So there is a huge possibility of measuring several different behaviours that have to do with society and its systems in society, enabled by the availability of data sources. This is equivalent to switching on a ‘telescope’ pointed at ourselves. Commercial actors can close the entire chain from the data to the value they extract, but these assets should be created and shared, so I am happy to see that there are provisions in the documents you have shared with us on using regulation and legislative pressure to actually create these assets. These assets are valuable at the level of an entire ecosystem. So, the logic that we need to imbue into this is enabling an entire ecosystem by creating shared data assets which represent a shared view of reality, available in real time and at a high resolution.

We also need to avoid the emergence of a gap in intelligence capability, in insight on reality, between commercial actors that create and manipulate these data, and public authorities and agencies that are tasked with managing the reality described by the data. It is true that data markets will probably partially bridge this gap, probably. Industrial data from data markets will generate some of this value. But I believe that we need to give higher priority to
addressing this informational asymmetry, and this justifies, I believe, specific regulation targeting non-personal data, and the use of non-personal data for public interest and for scientific research funded by the public.

Here in Europe we are quite ahead on the research side; when we speak of computational social science, network science and everything that has to do pre-competitively with the modelling and forecasting of behaviours at an aggregated level. This research was supported extensively in Horizon 2020 and FP 7, and I think we should capitalise on this in bridging the aforementioned gap.

About the difficult line that might be drawn between personal and non-personal data, it is true that there are always risks of re-identification, and as we become better and better with algorithms these risks actually increase, but I think we have to take a proactive approach there and actually invest into developing further technologies that can help us have guarantees on the risks of re-identification. In particular, there is research on blockchain technologies, distributed ledgers, the use of homomorphic encryptions, the use of dynamically aggregated or, adaptively aggregated data, the use of surrogate data, synthetic data modelled after the original data. There is a portfolio of technologies and technical possibilities that we can and should use. We should foster reflection on adopting these technologies, so that we can effectively use data science and unleash the public value it can generate downstream.

ANN METTLER

Excellent, thank you so much and the last one in this round is Professor Leiponen, please.

AIJA LEIPONEN

Thank you. I will focus on the creation of the industrial Internet of Things. I hear my colleagues have already addressed already a lot of other issues, and those are important too, but I'm just going to be focusing on this one. If we focus on non-personal data, much of the industrial Internet data maybe non-personal. If you think of let's say manufacturing operational data, production data, or logistical data in the Internet of Things. There is a lot of it. How valuable is it? It's probably valuable to the organisations that collect it and analyse it. How valuable it is outside those organisations is not known because we haven't done that much and this is work in progress in a lot of industrial Research and Development projects.

It seems that potential data sharing arrangements might be created through some kinds of data pools. This is perhaps analogous to patent pools which would include a consortia of firms, that after some multilateral contractual arrangements share their operational information. Having talked to a lot of companies and scientists involved in these initiatives, it is clear that we don't yet understand the competitive implications of this, which would need to be studied. If industrial players set up data pools, what are the competition implications? From a more
practical perspective, when companies set these up, what they are wondering about is what happens, even if they manage to set up the consortium and write the contracts, what happens to the data that has already been shared if the consortium breaks apart? There is uncertainty about the rights to control the data, beyond the functioning consortium, when there is a contractual breach, and how to prevent third parties from using the data.

There are problems also in just setting up those contractual arrangements. How to write it up in the first place? There is a big learning curve there. Companies do not know how to go about doing that and contractual templates might actually be helpful. Reasonable contractual practices, and also reasonable monitoring and auditing practices to track where the data is being used in that consortium, might also help companies get over the hurdle of setting up such arrangements.

Jumping into the General Data Protection Regulation, one of the questions was whether it makes a difference, and I do think it does make a big difference. From an economic perspective it’s very costly. It is a very costly piece of legislation. There is a lot of implementation cost for companies dealing with personal data, or data that can be viewed as personal data. And it will probably influence innovation, probably encourage innovation in certain directions. Probably towards data security applications and away from personal data services. That has a long-term dynamic implication.

One example that I heard recently from the Finnish context concerned a telecommunication equipment firm that requested to licence data from a telecommunication network operator firm to train their algorithms related to a set of services they would like to offer. The telecom operator dealing with personal data was not comfortable selling or licencing those data to the equipment developer because there is uncertainty or they did not know how that personal aspect of the data will be viewed legally, so they declined that deal. Ex-post attempting to obtain specific consent from individuals in that dataset would have been impossible. So, innovation in this case is prevented by the General Data Protection Regulation. Whether that’s good or bad, whether there could have been some potential harm that might have resulted from that Research and Development project, is an open question. Somebody would need to look into that, but it doesn’t seem obvious to me. On the other hand, there might be growing demand in other parts of the world for privacy regulation and privacy technologies in which case the European providers of those security technologies might actually benefit from that, and try to become leading providers of those technologies and services.

Just one word about localisation restrictions, which I think are going to be rather futile. I don’t think they provide a lot of protection, but in some cases, super sensitive data might benefit from that extra protection, but they would need to be studied case by case to justify for restrictions.

MARIO MARINIELLO

Thank, thank you very much. I will now read question four as with the previous question, you will have 5 minutes to reply. So:
The Commission intends to engage in dialogue with stakeholders to improve the EU framework for data access. The following goals are pursued: improve access to anonymous machine-generated data, facilitate and incentivise the sharing of such data, protect investments and assets, avoid disclosure of confidential data, minimise lock-in effects. Do you agree with these goals and what would be in your view the best way to pursue them? Which ones should be prioritised, and if any, what other objectives should the European Commission aim to achieve and why?

So this time we start with Professor Stucke. Here you go.

MAURICE E. STUCKE

Ah, yes, if I could take at least one or two additional minutes for this question and take that from question number five.

I agree with the Commission’s goals with a couple of caveats. One caveat is that the goals should be broader than minimising lock-in effects. The goal should be to minimise anti-competitive distortions in the marketplace. This broader goal would address additional barriers to the flow of data. One potential barrier that I mentioned previously are these data-driven network effects. Data-driven network effects are not necessarily bad. In fact, users’ utility increases as other people use the product. But with these data-driven network effects, strong firms can become even more powerful until they dominate the industry. This area is tricky because you can’t fault a firm for getting larger because of these network effects, but you still want to explore how you can promote competition and the free flow of data in markets with these data-driven network effects. This will be particularly important with the rise of these digital personal assistants. You may see them in commercials already. Amazon is offering, Alexa, Google is offering, Home. Since these digital butlers will be a key gatekeeper of the data collected from the smart technologies in our home, one concern is the super platforms’ abusing their dominant position in limiting access to this data.

You also want to examine anti-competitive distortions in the marketplace, as I mentioned earlier. One area would be data-driven mergers and in particular, vertical mergers, where let’s say the largest user of a particular type of data acquires a leading supplier of data. You also want to look at mergers that fall outside the traditional paradigm of competition policy. Suppose Google if they were to acquire Twitter. That wouldn’t historically be a horizontal merger, as the companies do not directly compete; nor is it a vertical merger, nor is it a conglomerate merger. Nonetheless, these types of data-driven mergers can have a negative impact on the free flow of data.

Another area you want to consider are abuses by dominant firms and these abuses can take various forms. One would be exclusive dealings to prevent rivals from accessing critical data, second would be exclusionary practices that prevent rivals from achieving scale and thereby collecting data. Third would be dominant firms leveraging their data advantage in a regulated
industry to another market. Fourth would be dominant firms increasing their customers’ switching costs. In order to maintain its data advantage and prevent rivals from achieving scale, a monopoly can make it harder for its customers to switch. If customers are then locked-on, locked-in rather, the monopoly can continue to acquire the data, and maintain its power. The General Data Protection Regulation helps address this but you still have the concern of dominant firms using other tactics to make it harder for customers to switch, which data portability won’t necessarily remedy.

And the fifth area of potential abuse would be vertical integration by a dominant platform operator such as when Google vertically integrates and starts competing and they have a ‘frenemy’ relationship with these apps.

My second caveat is that the ultimate aim is not to improve the free flow of data per se but to improve overall welfare; so you also have to consider any potential anti-competitive risks in increasing the free flow of data. One concern is that promoting the free flow of personal information can facilitate price discrimination. Another concern that we explore in our book ‘Virtual Competition’ is how increasing the free flow of ordinary market data in some industries can facilitate tacit collusion. And I’m not talking here about sensitive internal business records. Rather, tacit collusion is fostered by increased market transparency, generated by the free flow of ordinary data collected by the Internet of Things, and artificial intelligence. Companies then can see what their rivals are doing, they can also see what customers are doing. In some markets, this increase in market transparency can foster tacit collusion. The important thing here is that tacit collusion is beyond the reach of EU and US competition law, but the outcome is bad. Namely, consumers end up paying more or getting less than they would otherwise get in a competitive market. So you want to ensure that the free flow of data ultimately promotes welfare, and that the company’s interests in collecting and using data are aligned with society’s interest. Thank you.

MARIO MARINIELLO

Thank you very much. Professor Swire.

PETER SWIRE

Yes, thank you. And first I’d like to say that the discussion of competition law that we just heard was I think very far more sophisticated in discussing issues related to portability, than the discussions that I was able to find, at least in public, in connection to the General Data Protection Regulation data portability provision. The next question is more about portability, but the comments we just heard are a much fuller anti-trust explanation of what’s relevant than I have seen previously, and I really appreciate those remarks.
In terms of question four, and I'll turn to portability more a little bit later, a first point is that **removing barriers to sharing information doesn't mean there will be sharing of information.** We've seen this in cyber-security, where the United States has gone through rounds of efforts to eliminate barriers for sharing for cybersecurity purposes. If we share for cybersecurity purposes, that can be helpful because we can spot the bad guys who are attacking us. But, you not only have to get rid of barriers but you have to have some incentive to share, and often self-interest means that a company doesn't find any reason to share the information. So you can't just think that magically sharing will happen if barriers are removed. You're going to have to look at the incentives of each player to see what they are going to do with it.

The second point that I'd like to say is that in the Staff report and the other reports, I felt that there was a tension between two different views that maybe haven't fully surfaced, and one view you might call the intellectual property side of data, which is how do we ensure that companies get their rewards for their investment, and that can be a sui generis database protection, or it can be trade secret protection. And in that view the idea is we want to have companies getting proprietary yields from when they invest in data. But there is another view which is quite different which is that the more open, the better, which is going to be that we think it's going to be the best outcome for society if in general there is going to be more data in the data pool for everyone to play with. And I didn't see a very clear explanation of when each of those goals would apply. In the abstract each of those sounds good; openness is good, and also reaping the rewards of your investment is good. So I can't resolve the answer as to when each is better but I think, I sound like a Professor at this point, more research is needed to delineate when the intellectual property approach is more important or when the data pooling is more important.

One possibility where data pooling is important is to consider that essentially you're creating public records. Records that are going to be available for everyone in the public. There has been quite a gap historically, between public records in the United States and public records in many countries in Europe. The US has leaned towards having more information in public. You can find out how much my home was sold for, and who holds my mortgage in the United States. In many European countries that wouldn't be public. And, so one approach when you think you want to have more data be open is to explicitly decide that some category of data is public data and at that point, privacy rules wouldn't apply because it is open. And some countries in Europe have broader public record rules. Sweden does historically around income and various other things. So, I think those are the points for here, I think I'll come back to data portability later but I do want to appreciate the remarks we just heard about the wider range of anti-competitive practices to have concerns about here than just lock-in. Thanks.

**MARIO MARINIELLO**

Thank you very much. Professor Wendehorst.
CHRISTIANE WENDEHORST

Thank you. The empirical data we have seem to suggest that there is a tendency that data are kept within the company and are not shared with others and this is seen as being a problem.

Let me make three remarks on this. My first remark is this: you can't have your cake and eat it. When we discuss data protection, we say that keeping the data within the company is what we want, what is precisely the ideal. We want to have a clearly defined purpose and want the data to be in one place, and as a matter of principle we do not so much like them to be passed on, and passing them on needs justification. So there are two potentially conflicting goals, and we have to reconcile them.

Second remark, I think competition law is really the area to deal with the emerging issues. Building up monopolies is not a new phenomenon. Vertical integration is not a new phenomenon. We've had that for decades, if not centuries, also outside the data economy and we know how to deal with such developments. It may at times be difficult, but we know in principle how to deal with it. And there are some court decisions like IMS Health, Magill, Huawei etc., which show that in principle, competition law also works in the data environment. We may have to consider some changes, for example when it comes to merger control it may be not sufficient to only look at turnover figures, so we may have to make some adjustments here and there, but in principle I think competition law is the key to our solution.

Having said this, there are certainly some additional measures which I would like to recommend. One is the development of standard contracts, standard licences, with guides on how to use them. This would not be coercive, it would just be something that would be offered to businesses in Europe and they can make use of it or not which would be beneficial in particular for SMEs. Then I could imagine targeted harmonisation of data contract law, clarifying the role that is played, for example, by property law and possibly introducing some sharing obligations for data analytics carried out in the name of the public interest, a little bit like in the 2016 copyright proposal, but of course also different.

Just my third and last remark. I am very sceptical when it comes to introducing something like a data property right at this point. I think this is definitely immature as it might have a disruptive effect on the data economy, and it would be difficult to control and to define. It might achieve just the opposite of what we want to achieve. Thank you.

MARIO MARINIELLO

Thank you very much. Doctor Cattuto.
Ciro Cattuto

Yeah, thanks. Overall I agree with the goals to improve access to data. Currently, especially for research and for the value it can generate downstream, the barriers seem to have to do with uncertainties about the liabilities in sharing data; the costs of post-processing data to make it available to researchers; or when the data cannot be moved, the risk of giving third parties access to one’s secure infrastructure.

In general, on creating this compositionality that unleashes value from data, there is a general perceived imbalance from the commercial sector between the risks and the benefits of sharing data, and I think the imbalance is real. There is uncertainty. There is also, as it was pointed out by the colleague, a general lack of standard contracts that can be used as blueprints for setting up data-sharing agreements. Right now data sharing happens more often than not in a point-to-point fashion, and this leads to delays and extra efforts, that would be avoided if we had blueprints for this kind of agreement. Moreover, point-to-point arrangements tend to discourage replication which is a huge problem for research because you end up with point-to-point relations and generally one-off results than the community cannot replicate. And this is a recipe for bad quality science - lack of replicability overall.

So in general, anything that we can do in order to lower these barriers for research and for commercial actors alike will be valuable. There should be, I believe, a stronger focus on improved access to machine-generated data to support the excellence of European science.

European science has actually moved fast on the underpinning knowledge needed to extract value from data and now it needs to be empowered with the right level of data access. I would like to call for a sort of ‘Big Science’ vision for European data science. There is already a strategy fleshed out for communication networks and high-performance computing, that is the foundational layer. I think it would be interesting and important to flesh out the strategy for a fully European ambition of research enabled by data that might be commercially held at the origin. These data, as mentioned above, provide information about processes that are core to the functioning of civil society, so it is important that we create these value chains.

One of the questions was whether there is a gap between the private and social value of data owned by private firms. In respect to social welfare I believe this gap is actually there and should be addressed. Incentives will go a long way, but I think that it would be very valuable to have a library of collaboration patterns around data. There are experiments and projects along these lines in the US, one in particular comes to my mind, the Data Collaboratives project by the New York University Governance laboratory, GovLab. What they do, which I find very useful, and could be replicated, is to create a library of success cases where data were shared between public stakeholders, commercial actors, government, non-profits, etc. Such a library of success stories could, on the one hand, inform policy making by pointing out what works, what does not, what are the friction points, whether there are some regularities there that could be
captured and turned into policy. On the other hand, it would generate awareness about blind spots. By mapping out data sharing exercises, it might be possible to see that some value we expect to be generated is not generated, and this can pose targeted questions and lead to a call for action.

In general I find that in this discussion about the data ecosystem there is a **blind spot about the potential use of data by philanthropies and foundations**. Europe has got a very rich ecosystem of foundations. Next month, at the annual general assembly of the European Foundation Centre there will be a session about data science for philanthropy, which is not just about reasoning on how to evaluate the impact of philanthropic actions by using data, but it’s also an opportunity to be proactive in funding interventions that bring together different types of actors with the goal of sharing data. This is another way to incentivise, for public interest, interactions and data exchanges which otherwise the market would not generate. Thank you.

MARIO MARINIELLO

Thank you. Professor Leiponen.

AIJA LEIPONEN

Thank you. I will jump right into the discussion of market, potential market failures in data. **I think we can easily imagine that there will be market failures in data, but we don’t know where they will be.** So, because of the non-rivalrous nature of data, there is, with very high likelihood, opportunities to use the same data elsewhere in the economy. But those trades might not happen because of uncertainties in the marketplace and difficulties of knowing what the potential uses might be. **And ex-ante regulation for those eventualities would be very risky and would be difficult to see how that could be done.**

On a general level I **can see potential for incentivising firms to share their data through some kind of Fair, Reasonable And Non-Discriminatory licencing or some other mechanisms** when that’s associated with, for example, Research and Development investment subsidies. So if there is a research programme, a Research and Development programme, that is incentivising technology development in a particular area that might be combined with some expectations for sharing data that is being created as a side-product of that Research and Development. **But if there is aggressive legislation to share data that private actors already hold, that creates an incentive to not continue to hold those data.** Data sharing requirements can backfire, and we would need to know when they do so; when we should expect firms would prefer to get rid of their data rather than share it.

**The value of data, as that of other forms of intellectual property, is largely determined by the context in which it is used.** Therefore there is unlikely to be an open market and prices for data, in any meaningful way. Unfortunately I think these kinds of markets will
be riddled with and implemented with price discrimination - that will be an inherent part of market formation. If most trades will be bilaterally negotiated, there will almost by definition be price discrimination.

We’re dealing with markets where there is probably going to be large fixed costs to create the data assets and low or zero marginal cost, and so we cannot fall back on the very traditional anti-trust analyses to deal with this market. More likely we will see price discrimination strategies for companies creating data assets trying to find who is willing to pay for these assets or services.

I would also note that data is usually an intermediate input; it’s not a final output, so it goes through a production process to create more value out of it. You mentioned data value chains, and that’s an important perspective into understanding how value is created in the data economy. There can be many steps, and the original data resource can be manipulated many times in different ways, and subsequent outcomes can be again manipulated in other ways for many potential markets, and so, that’s just the nature of this input, and the nature of the asset. We have to keep that in mind when we think about markets for data.

Some specific initiatives that I’ve seen in those communications from the Commission; I’ve seen producer rights mentioned as a potential approach to strengthen the data holders’ rights in commercialising their data. In some cases I can see that might enhance the benefit of sharing their data when there is a reason to engage in sharing or selling data, or licensing data, but the data holder is concerned about incomplete contracts, including third-party implications, and long-term implications. But I would also be very concerned if such producer rights were associated with the ability to block competitors who independently create similar datasets and then are not allowed to commercialise those. And I don’t know how to deal with that problem, legally.

Another initiative I have seen is the rights of users, especially device owners to their own user data. Who should have rights to that? Is it the manufacturer, or is it the user? Device owners themselves have rights to that. And there are likely to be innovation implications associated with that decision. Incentivising the owner of the device to utilise their own data and potentially share it with third parties might enable them to enter into that industry. I believe this is a situation where they may potentially be reasons to share those data.

One last thing I want to mention is that distributed ledger technologies, blockchains and such, might facilitate some of these rights issues in the future. Probably not in every situation and every industry and every case, but these technologies are worth considering in the data market. The Commission could find ways to support the development and application of such technologies in the data markets.

Thank you very much. So our last speaker, Professor Spindler.
GERALD SPINDLER

Thank you, and I can easily join confirming what has been said before, but I will just try to add some aspects to that. I think there are two flip-sides of the same coin, as Christiane Wendehorst has already pointed out, we are confronted with a question of access or use on one side and protection on the other side, which is mentioned here in your core question.

First let's have a look at the acts, at the use side concerning intellectual property rights. As I already mentioned, there is already in the Parliament the proposal on text and data mining of the Commission and which is really crucial for anything to make use of already existing data. If it's really true that this is a limitation to cover use of existing data and texts then we should affirm that the mere use of data does not infringe intellectual property rights; otherwise it would have a huge impact of anything concerning algorithmic and data etc. because then you need a licence which the publishers are already trying to invoke.

Secondly, Database Directive, and here we are confronted with the problem of what is protected in the Database Directive. Usually only the structure of the database is protected, but also collecting the data if there is a substantial investment. So what is then collecting, according to the European Court of Justice? It is not only about collecting existing data; it's also about adding something substantial. So how about now sensors which are collecting data; is that protected or not? So, we have to clarify that and it could be easily done in just amending the Database Directive which also needs to be amended, concerning other issues, like access to data.

Thirdly, we have the Know-how Directive which could affect the notion of industrial data here, but it is not, as somebody said, an intellectual property right, we really have to be careful here. It's not tradable in the traditional sense of intellectual property rights.

So there are a lot of legal uncertainties still there in the room which could easily be solved in the next coming years without stirring up a debate which would lead to the very bottom line.

Secondly, concerning more general ownership of data, we shouldn't introduce anything like that. It has already been mentioned by my colleagues from the economics side that it is very hard to here create property rights in an economic sense, in strict delineating the border lines, or what about similar datasets etc., as Christiane Wendehorst already really pointed out that this could blur all the lines between intellectual property rights and anything else. For me, the basic question is: are we talking mainly about business to business contracts, is there really a market failure? Are industries in other countries which also do not know what property right on data, are they failing completely? I haven't seen that until now; as an economist I would say that there is a prima facie proof that obviously the markets are working, somehow.
So next point then is antitrust law, and I mostly agree with my colleagues, this is one of the core issues; however I am a little bit more sceptical. If we take a closer look to the so called more economic approach in anti-trust law, it shows us that most of the cases are pending for years and years and this may lead to the very problem that anti-trust law may step into too late. Let us, imagine a Small Medium Enterprise trying to fight against one of the dominant market player on a private legal basis then it will easily end up two three years later, with the absolutely bankruptcy of the Small Medium Enterprise. Of course there are some prominent cases in anti-trust law, but usually anti-trust law steps in too late. It's an ex-post solution which may not in an economic sense really work out.

So, moreover, concerning the economic effects, I just wanted to call in mind that we are here faced not with a static competition issue, but a dynamic competition, and this is really very hard to assess how dynamic competition inter-temporal allocation of resources as has been called in economics, can be here assessed. We are confronted here with very dynamic business models which vanished over time; think of the old Microsoft debate in the nineties and this is not on the table anymore, concerning that.

And this leads me to the next point, what is the definition of markets, what is the dominant market player? We have learned a lot and sure we have to take into account dominancy concerning data and some, but I think we still need a lot of research also from the economic part of, to assess that.

So what would then be the solution to my mind? It could be thought of an introduction of something like in the Software Directive as already mentioned, or in the Database Directive, such as an extended right to have an interface, a right to have access to the data, but which is then as some sort of a negative part of 'property rights', not a tradable right in the sense of intellectual property rights but to have the right to access like in the software to the code in order to establish the secondary market. And it could be combined with Fair, Reasonable And Non-Discriminatory licenses for example.

Last point, concerning the platforms, it is absolutely interesting to see that there are no platforms obviously really working now, and we are facing this data now for more than ten years. Then the question surely arise of what are the reasons perhaps, if there is a market failure, what are the potential reasons for that. You named a lot concerning the platforms or patent pools etc. but it is not about the comparison with patent pools because these are referring to real property rights. We should more look into pools of know-how, sharing know-how, this could be combined with post-contractual obligations and guarantees etc. And there, the Commission could play a role in establishing standards and standard contracts for them, such blueprints contracts in order to establish these kinds of platforms which then can be to overcome the market failure.

Last but not least, and also referring to what Christiane already said, I'm in favour to extend the unfair terms and conditions directive concerning Small Medium Enterprises, in particular
by introducing on the blacklist something about data licence agreements and intellectual property licence agreements in order to overcome market failures. Thank you.

ANN METTLER

Thank you so much. The fifth question deals with non-personal data portability and inter-operability. You will have five minutes to answer. The question goes:

The European Commission intends to address issues of portability and inter-operability for non-personal data. Do you support that initiative and why? In your view what are the best ways forward to facilitate switching and to prevent lock-in while minimising the risk of undermining investments in the data value chain?

So we'll start off with Professor Swire please.

PETER SWIRE

Thank you very much and as I mentioned earlier I wrote a paper on data portability as proposed for the General Data Protection Regulation, and much of the analysis would apply here to non-personal data. So, before getting into the anti-trust points I made there, the idea that we should address market failures and not others, that’s something that I would take. The fact that anti-trust cases happen too slowly and ex-post, I agree with that, and I agree with many of the points Professor Stucke made. However, the anti-trust analysis I did to the right of data portability was much more sceptical as an anti-trust measure. So, if you look at anti-competition law in the EU, trying to help out consumer welfare, there’s at least three ways that the right to data portability in the General Data Protection Regulation departs a great deal from EU competition law. The first is that it applies to small, medium and moderate sized enterprises in addition to dominant firms. So, at least as I read it, if you are two or three people writing a software app, they would have to write portability in from the start. There’s not any real plausible anti-trust case that they are locking in or whatever. So attention to dominant firms is what people are talking about, but the rule applies across the board even to Small and Medium Enterprises. That seems over broad, it discourages investment in small firms that don’t have time to go write extra software and, and there’s often difficult interoperability problems when you write software. So the over application of it even to small firms is the first point.

The second point is that it really makes the rules about lack of portability into a per se violation instead of the rule of reason approach which is usually taken for exclusionary practices. We have heard reasons why refusal of supply or denial of access or whatever might be there for dominant firms, but there’s a lot of possible efficiency reasons not share data. One that gets used very often is to say ‘I’m not going to share my data with you for cybersecurity or privacy reasons, because I think it would be a risk in those ways’. Sometimes those are a pretext, it’s not really a cybersecurity argument, it’s an anti-trust exclusion problem. But sometimes the
cybersecurity argument is a good one about not sharing the data. And so, a per se rule, instead of a rule of reason, seems very different from European competition law. I think I combined my second and third points.

The other things I’d say is that in the questions there’s discussion about promoting standards for interoperability and standard formats. I think that there is a role for public policy and standards bodies and it can increase interoperability and create benefits. But I, would caution, and come back to this point, of not requiring non-dominant firms to have a regulatory burden out of all this. If the companies comply really with that it could be really difficult and expensive to do the software, and that’s the point that doesn’t seem to be widely discussed, as the right to data portability was considered, for General Data Protection Regulation.

Thanks very much.

ANN METTLER

Excellent. Thank you so much. Next is Professor Wendehorst please.

CHRISTIANE WENDEHORST

Thank you very much. When it comes to portability of non-personal data I would like to differentiate between two scenarios which I think are often confused. The first scenario is a contractual scenario in which data are held by contracting partners, for example a cloud service provider or they are held by the producer of goods or of digital content which a person has acquired, such as by a car manufacturer, or by another business that cooperates with the contracting partner or the producer.

The person running the car, or the person using the cloud service, has a contractually protected interest in those data, e.g. to get their e-mail back when they want to switch the provider, to get customer or financial data required for running a business back, including when the person is not satisfied with the service, and so on. In these cases portability and interoperability are crucial and must be provided for, and they would in many countries already be protected by contract law - for many reasons, including facilitating switching. They would even go beyond what we have in Article 20 General Data Protection Regulation because they would not be just for the raw data, (c.f. what the Article 29 Group has just clarified), but also for refined data where this is what is required. I believe we need something for this contractual scenario if there are indeed problems in practice, which I appreciate there are. I can repeat what Gerald Spindler has said: we need unfair contract terms control, we need lists of unfair contract terms that are specifically addressing data issues and we may need new contract rules that work for multilateral environments and that take data issues into account. So where data are required by a person to get what that person was entitled to expect under a contract, interoperability and portability - no matter whether data are personal or non-personal, no matter whether the person entitled is a business or a consumer - are absolutely crucial and must be addressed.
But then there is the second scenario, where data are collected solely for some other than a contractually protected purpose, e.g. the manufacturer of a machine that has nothing to do with weather conditions collects weather data, and the only interest, which the person owning the machine could have in getting the data back would be to use them as a bargaining chip for the potential next provider. In this case, I think the situation is very different, and we need a really strong justification for going into the direction of an equivalent to Article 20 General Data Protection Regulation, in this second scenario. Of course, as I have recommended data trusteeship I should have some sympathy for an equivalent to Article 20 General Data Protection Regulation because that would allow for a uniform approach to personal and non-personal data, but, from another point of view, I think there would be serious side effects. It might discourage businesses from creating innovative collections of data, it might endanger investment, and as Professor Swire has pointed out, it might be a disproportionate burden on Small Medium Enterprises. So in the second scenario I do currently not see a sufficiently strong case for going into the direction of an equivalent to Article 20 General Data Protection Regulation. Thank you very much.

ANN METTLER

Thank you. Next up is Doctor Cattuto please.

CIRO CATTUTO

Thank you. I will just add a quick comment, since most of the comments I had have been pretty much covered by what has been said until now, in particular by Professor Swire. Just one comment on standards: achieving standards is certainly something that the EU has to work towards, also as a way to level the ground for competition, but in dealing with global players, imposing standards and making them binding will lead, I think, to a protectionist approach to standards, and eventually, I think, it might incur the risk of raising barriers for the entry into the market of smaller players of SMEs in particular.

Standards, especially in the Internet of Things domain, will emerge out of the interaction of market players. Most of these players, whether we like it or not, will be global players, they will not be European players, and probably they are already dominant players. And they have all the technical and pragmatic means of making their standards successful. So I think that engaging in dialogue is important, but relying on binding standards will just alienate opportunities for our market and will create a barrier to entry into the market.

The other challenge here, especially in the Internet of Things domain, is that data portability meant in terms of provisions to extract, move and import into other systems, data generated by sensors can be technically very difficult to achieve. This might impose a heavy burden on the data generator, or it might expose, on sharing the data, competitive
intellectual property assets of the data generator. We are reasoning about systems where
the sensor will be just one part of the valuechain, but the sensor itself might incorporate a
significant level of intelligence, and might be endowed with technical capabilities that are advanced:
sharing the raw data it generates might significantly expose a lot of the intellectual property and
danger investment on part of the data generator. I think this is a huge challenge and this just
calls for more research in this direction. Thanks.

ANN METTLER

Excellent, thank you so much. Next is Professor Leiponen please.

AIJA LEIPONEN

Okay, thank you. I’m trying to focus on switching costs and lock-in in non-personal data settings and
in particular, within industrial settings. And it does not seem to me that switching costs are
a huge issue in that space. The portability of data from one industrial setting to the other is
probably not what is driving the decisions or influencing the competitive outcomes in that area, so...
Personal data switching costs may be a whole different issue but in the industrial setting I don’t see
that as quite a central influence in the competitive outcomes.

There was a question about welfare effects related to data portability on multisided markets. It is
well known that multisided markets may concentrate market power very substantially, but
sometimes they’re also the only way to realise network effects; and these network economies are
quite central to communication networks. And so we need to deal with the market power as it
arises with the network effects, but it is not clear to me that portability will solve that
problem of network effects increasing market power for platform providers.

Data standards I think are a big issue. This seems a very mundane and technical issue but it is
actually difficult to address in many industrial sectors. Standards appear to be quite
fragmented across industry verticals and even across organisations, as large organisations
may have their own legacy formats and ways of processing and storing data. And for that reason I
would be, I would encourage any efforts to create open European or global standards
around data perhaps for storage and exchange. One possible area where I see that to go
forward with that is applying, developing such standards and applying them for public sector data
within the European Union and making public sector data available according to those open
standards. This might generate use and innovation around the data and adoption of those
standards and formats at the same time. Open and standardised public data would potentially
enable innovative start-ups in using both the standards and the data to enter the data economy.
And one interesting part of that picture is the 5G, so called 5G standards that are more at the
network level may also play a role. There’s also some fragmentation of the network standards
themselves. And I’m not sure how to go about harmonising all that, creating one big Internet of Things, but fragmentation is always a problematic in communication networks.

ANN METTLER

Very good. Thank you, next Professor Spindler please.

GERALD SPINDLER

I think I can be brief because I mostly agree, let’s say, to what my speakers before had said. First of all I’m very sceptical about any welfare effects of some sort of standardisation, because usually, if there is really a need for standardisation, the industry will call for it. They will just give a mandate to the standardisation organisations and mostly they do not need some support or stirring by the Commission. Even though you would identify some form of market failure, then it would be a question why shouldn’t we mandate the European Committee for Standardisation (CEN) for example with these standards, so that industry can go for it.

And I just wanted to pick up what my Finnish colleague already said and she was absolutely right: we have to do our homework concerning standards in the states and I can just tell you the mess. We have in German administrations that some agency cannot communicate with some other agency because they are applying for example different standards. But how to overcome that problem? That would be highly difficult for the European Commission, as you won’t tell German administration which standards they have to apply! But this is just a factual problem which we are facing there. And so it really would be to say bluntly it is quite a joke if we are now trying to establish standards for industry and we are not able to do it, at home, even in our own state at the federal level.

So I have my doubts if efforts in standardisation are really necessary, in particular by State intervention. If so, standards also should be related to Information Technology security questions. If we are going to establish standards for data which are to be transmitted (but also of course for software etc.). We have to care for security standards as well. So standards yes, but in other areas than here, because usually platforms (for example in industry), are creating them themselves. I’ll just call to your mind, for example, the banking networks which established over decades these interoperability standards for data between, because they had a need for that. Exchanging financial data, they established worldwide networks, creating a standard for that. So I do not see really here a need, and in contrast, there could be detrimental effects to the welfare. Take for example trading platforms with reputation systems; if you just make it all interoperable as far as people from E-bay – I have to believe it, I cannot verify it but it’s true then their asset – their core asset in the networking platform is their reputation system. So if you transfer that, to other systems, to other platforms then they would lose a lot of their investment there. If it is true that statement, I just cite it here. Thank you.
ANN METTLER

Very good. Thank you so much, and last Professor Stucke please.

MAURICE E. STUCKE

Okay. To the extent that data-driven network effects and market power are at play, increasing portability may lower switching costs and entry barriers, and also reduce quality differences among products. But, like Peter I would encourage you to look beyond data portability. Promoting data portability will not necessarily remedy every anti-competitive distortion in the market place, and one may need other more finely tuned measures to promote the free flow of data.

One example is anticompetitive scraping. One complaint now before the European Commission against Google is that it scrapes content from rivals and posts that content on its own websites. Consumers, as a result, remain on Google’s websites and Google collects the consumer data. Now, allowing consumers to port their data won’t necessarily prevent or remedy this anticompetitive scraping, which adversely affects companies upstream. Another example involves one of the four V’s of Big Data, namely velocity; one illustration is the real time geo-location data for turn-by-turn navigation apps such as Google’s and Waze’s navigation apps. The velocity in collecting and processing data is key. Even if you allow consumers to later port their geolocation data, that won’t be of much help to rival navigation app providers. To be competitive in some markets, rivals may need access to that geo-location data at the same time.

So one thing you may want to consider in some of these markets where velocity is key is to shift from an ex-post to an ex-ante framework, whereby for example, individuals can elect ex-ante the simultaneous collection of their data from their own data locker. And I would say, the broader point, to follow what Gerald Spindler mentioned, is that anti-trust won’t always be a good solution, particularly from an ex-post perspective. What is required is greater coordination among competition, privacy, and consumer protection officials to identify the necessary preconditions for both privacy competition as well as a competitive data-driven economy overall. Thank you.

MARIO MARINIIELLO

All right [cough], thank you very much. So this is our last question, to gather your bottom lines, and each speaker will now have just one minute to reply. So:
'In a nutshell, what is your main message to the European Commission, regarding what should or should not be done about the data economy?'

Please.

CHRISTIANE WENDEHORST

Thank you, well one minute, that's difficult. I think this session has shown the importance of interdisciplinary research. We need to know about the technical possibilities, we need to know the economic impact, we need to have legal perspective and if in doubt we should probably take a cautious approach: start with minimum invasive measures and see what the effects are and not rush things with something the effects of which we cannot foresee.

Thank you very much.

MARIO MARINIELLO

Twenty seconds! [chuckle]. Doctor Cattuto, please.

CIRO CATTUTO

The data revolution hinges critically on data reuse for purposes not anticipated originally, so we need additional measures to maximise data use and reuse for public interest and to level the ground between public actors and commercial stakeholders in terms of the intelligence and decision-making capabilities that might have public interest. And, of course, we need to achieve this while, at the same time, protecting the investment of commercial stakeholders.

On the science side, we have an opportunity to improve the competitiveness of European science by introducing incentives and legislation for accessing non-personal commercially held data for fundamental and applied research in a number of interdisciplinary research domains.

Finally I believe that we need a big science vision for European Data Science. Shared digital facilities for processing, cross-mining, analysing data, that can support the work of a broad interdisciplinary community to advance our scientific knowledge as well as to improve crucial functions of the EU. Thanks.
Thanks, Professor Leiponen.

Thank you. I have three points at this point to make. I think it's important at this point to understand the complementarities and the systemic nature of digital network data, software and algorithms, models and intelligence, and connecting the network itself. It's a very complex value network and if we don't understand where value is created, what are the drivers of investment and innovation around data in that value network, ex-ante regulation can backfire and destroy those incentives.

The key issue is encouraging European investment into software-based, data-driven services and perhaps products, and I would encourage policymakers to deal with market power issues later through competition regulation, rather than trying ex-ante to influence the sharing of the benefits of the data economy before the benefits have even been created.

Thank you very much. Professor Spindler.

Once again I can easily join the statements of Christiane Wendehorst and Professor Leiponen. First of all we need a step-by-step approach, not too invasive. This has to be flanked by a lot of empirical research as well as economic and informatics research.

Secondly, if we look at short-term and mid-term solutions, first in that which could be in the Parliament [outcome on] text and data mining should be carefully scrutinised, once again concerning for example commercial data mining. Then we have to look at the standard terms and conditions Directive adding here something to the blacklist. This goes as well for the unfair competition Directive. There you can easily add clauses to the blacklist.

Concerning antitrust, the Commission itself could do a little bit more concerning the definition of markets, concerning data without changing anything in the antitrust law.

Then thirdly but not least, it is creating or introducing some sort of interfaces like in the software Directive and the database Directive combined with the Fair, Reasonable And
Non-Discriminatory Licence argument there. And last, standards should be established - but this goes beyond what you are asking here - in IT liability and security, which really there has been, there is really need for doing something. Thank you.

MARIO MARINIELLO

Thanks. Professor Stucke.

MAURICE E. STUCKE

Yes, the aim here should be to develop an inclusive data-driven economy that benefits more than 1% of the population. One thing that we heard today is that you cannot assume that market forces alone will yield the benefits of the data-driven economy while mitigating the risks. Another thing that you heard today is that you cannot assume that one agency can do the job. Just as you need to break down the data silos and the geographic silos, you also need to break down, as Peter Swire mentioned, the silos of the governmental agencies. Here you need greater coordination among the privacy, consumer protection and competition authorities. The good news is that the efforts of the European Data Protection Supervisor in seeking to launch a digital clearing house for enforcement in the EU digital sector. That’s a positive step.

So in a nutshell, the goal for a data-driven economy should be an economy that’s inclusive, protects the privacy interests of its citizens, protects the citizens’ overall wellbeing, and also promotes a healthy democracy, because the interests here at stake go beyond our pocketbook. Thank you.

MARIO MARINIELLO

Thank you very much. Finally, Professor Swire.

PETER SWIRE

Well, thank you first of all for this outreach, and for your thoughtful process in writings that the group, the staff and others have done. My first point was going to be the importance of better coordination of the data economy, competition and data protection officials. And - we just heard that from Professor Stucke – I think getting concrete ways of achieving multiple goals is really important and the example from Finland is just one of, I think, probably many examples.
And it’s difficult because privacy has the fundamental rights status in Europe and so is very difficult almost to talk about any limits on that. But without that it will be very hard to achieve any progress on that here. And the reason is that there is a presumption against processing when it comes to personal data. And then the broad definition of personal data becoming increasingly broad means that almost anything can start to seem to be personal data, and that leads to a broad presumption against processing data for a wide range of settings, so that’s really a fundamental tension that will need to be resolved.

One thought on ways to perhaps address it is to think more about how to do sharing but under strict organisational controls, not with public posting of data, but perhaps in the Finland example there could have been a very clear contract that would be used for research purposes. Research is a word that many people favour for many good reasons, and having a more extensive set of organisational controls to permit research that will enable innovation might be one way to frame both data sharing and data protection goals. Thank you very much.

ANN METTLER

Excellent, thank you so much. This brings us to the end of this hearing. I want to warmly thank all of our experts, we covered a lot of ground. I think we learned very much. We are deeply grateful. I want to just quickly reiterate the process, which is that in the coming days, everything that was said here today will be transcribed and would then be submitted to the public consultation on data that is ongoing. Just to say, we had some very good colleagues around the table, they’ve been very patient. We’ll be serving coffee now and I would encourage our external experts as well as our colleagues to perhaps stick around for a few minutes because you will certainly have some questions of your own.

Before I let you go, there are two people I need to thank. It’s firstly Mario Mariniello, who organised all this and secondly, another colleague who was in the room earlier, but now I don’t see her. Her name is Cristina Ruiz and she did all the heavy lifting on logistics and putting everything together. A lot of work goes into it so I want to warmly thank those two colleagues and perhaps we give a round of applause.

[Applause]

Mostly, of course I want to thank our external guests, our experts. You’ve done a lot to enlighten us today and I can really only warmly thank you for your contributions, for making the effort of being here.

We will now serve coffee so it will be an opportunity to actually say goodbye to Professor Swire and Professor Stucke. So I’ll wave goodbye to you. But before you log off, please also a warm round of applause for our external experts.

[Applause]
Thank you so much, this concludes the hearing.

**The text reported herein has been obtained through manual transcription of an audio recording taken during the hearing. The text has been adapted with some stylistic corrections in order to facilitate the comprehension by readers, following the speakers’ feedback. No substantive addition or change that could affect the interpretation of the speakers’ statements has been introduced in the text. Although the EPSC believes the text to be most accurate, mistakes and omissions ought not to be ruled out.**