Report on the meeting of the Digital Lab workshop

Meeting report – May 5, 2015

Big Data: Beyond the innovation vs protection dilemma

Louise Decourcelle

Speakers:

- **Benjamin André,** CEO & Co-Founder of CozyCloud
- **Michal Boni,** MEP (EPP, PL), Founding Member of the Digital Agenda Intergroup
- **Eric Peters,** Deputy Head of Unit, Digital Single Market, DG CNECT

The digital identity of private and public European organizations represented 53 billion euros in 2011, according to the Boston Consulting Group. By 2020, all sectors combined could represent up to 330 billion. Data have become the new currency in the digital economy and the growth opportunities in the areas of Big Data or the Internet of Things are often branded as another Eldorado. But such wealth might dwindle if confident relations are not established. The American hegemony in this field seems almighty, but the PRISM scandal has shook the confidence of many Europeans. With the upcoming Digital Single Market Strategy, Confrontations Europe gathered stakeholders with a different background–from a startup to the European Parliament–for this timely workshop to discuss the following questions: What options does Europe have to reclaim control over our data? Should it reinforce its existing legislation through new fundamental rights, by new tools? How to concretely assure a just and fair balance between freedom of speech, private life and economic development of innovative activities in Europe? What actions could Europe support to develop common norms that would allow European start-ups to grow in a trustworthy climate, where the new uses would be appropriately monitored?

FULL REPORT

MEP Michal Boni (EPP, PL), funding member of the Digital Agenda Intergroup

He welcomed the initiative of the future DSM strategy. Before to get into the debate of innovation and protection dilemma, he made 3 general comments:

- Firstly, he argued that **freedom of speech, access to open sources, privacy, developing digital economy and using the data are not contradictory**. They have to go hand to hand. This is the necessary background for creative innovations. He stressed that processing the data can be very fruitful. Processing all kind of data creates the new economy engine and it is the core of new digital businesses.
- Secondly, he stressed the need to **create the good conditions for the trust & confidence as a background for the future digital economy.** How to build the confidence? He noted that we live in a time where we can experience monetization of data. Therefore, trust is the new background for future digital economy. It means that we need a balance between our citizens' – as users, as consumers– needs related to privacy and possibilities for business. We don't want to have a regulation out of reality.

Regulation should work and should be implemented. It's not an easy task so it's important to gather stakeholders to discuss it.

- **Building the rules supporting privacy all around the world**. The EU aims to harmonize. In this regard, he argued that European solution should be a reference point for global solution. The EU should work with the USA to solve problems of umbrella agreements and Safe Harbour agreement. But the EU should also support them in their ways to achieve their goals related to privacy protection. In the USA, they have prepared the privacy act: it's a different solution from the EU but really similar. He argued that EU regulation should be good point of reference.

Also, he emphasized the fact that we should understand that **processing of data is key for the development of digital economy**. One of the opportunities of processing data is related to personalization (i.e. if we can process the data, we can prepare for much more personalized services and goods). We also should remember there are many data: some is used by governments and public authorities and processing those data are really useful for the digital economy.

Regarding the innovation and protection dilemma in data protection, Michal Boni expressed 6 points:

1. **Need the data protection package**: it means not only to finalize the regulation, but also the directive and also review the e-privacy directive and have a privacy package (all those solutions are needed for us − as user, as consumer, as citizen− and for businesses also and for better preparation for governments for better administrative services). The Council is working and we hope they will finalize their position in the mid of June. The Member States are currently working on chapter 3 on data subject's rights. According to Michal Boni, the trialogues will be some kind of new start of debate. On one hand, he stressed that from a political point of view and from a Parliamentarian point of view, we should achieve of our goals but on the other hand, he stressed the need to understand some aspects of the member states during the trialogues. His idea is: if we want to have a good implementation, we should start to create a kind of round table with all the stakeholders. Thanks to his experience of being in ministers, Boni knows that the work is not finished when you have a regulation: the work should be started. There should be a social process during the trialogues and beyond. There is a need to consider many aspects of the implementation because of many differences in countries. For many countries: the implementation of this regulation will be completely new. So we should support all the member states in this process.
2. **Who should be the owner of the data**? This is not a philosophical problem: citizens should be the owners; this is the basic background for data protection regulation that somebody should ask us our agreement. On the other hand, we should also remember that there are some data for which the government or public authorities are the owners. And there are also some data produced and processed in the business. He stressed the need for clear rules related to the problem of who is the owner of the data. At the same time, we need clear rules about the problem of how to share the data? For the digital economy, sharing of the data under clear rules is very important. We should not kill the possibility of sharing the data. Need to address what kind of rules should be related to sharing the data?
3. **Infrastructural problem is important for the development of data-driven economy**. Issue of infrastructure: access to the internet, access to the network. Obviously, need to finish our work in Europe on 4G and start the work on 5G and we need fast internet. On the other hand, we need some kind of European network of high-performance computing infrastructures. This is important for processing the data. This is related to storages and to platforms. It's about creating new possibility for European networks rather than fighting American companies. Technologies such as e-health or m-health are really important for the future but need to be developed under strict conditions related to their sensitive data related to health care area. But before considering a

discussion on the implementation of e-health: need to address the issue of storage: where will be my data? It is important for communication to the public.

4. **Labour market and new skills**. We need not only skills related to processing the data but we need engineers, data analysts. We also need some kind of building awareness of privacy and data protection. He stressed the fact that people don't know what it means that business and public authorities are using our data. We should inform, build and increase awareness. He also presented the idea of introducing some kind of coaching, stressing that companies have a moral responsibility[1].
5. **Cooperation between all contributors to data-driven economy**. Need some kind of forums, exchanging views, stakeholders' cooperation and be open for many new things.
6. **Security:** Need stronger security of our data, of all processors. For instance, sharing economy requires better security. We are starting the discussion on security by design on our devices but 75% of mobile devices are produced in China so we should discuss about security for people (such as the encryption possibilities) and for networks. There will be a public hearing at the European Parliament on July on data-driven security. Many new technologies can allow us to analytical approach related not to personal data but on data processing and give us some kind of prevention.

Eric Peters, Deputy Head of Unit – Digital Single Market, DG CNECT

He started by stressing that the announcement of the Digital Single Market Strategy will be a major event. He shared the views expressed by Michal Boni saying that the digital single market goes beyond economic development, it's a societal issue. He stressed that this is a hard work because gathering different political views, and views from South and North, is very challenging. On the other hand, he pointed out that if the EU delivers the DSM it would be a good example of what EU can do best.

Regarding the DSM strategy− which was due to be released the next day−he gave us some personal thinking of where we stand. He made a general comment underlining that innovation and protection are not opposed. However, he acknowledged the difficulty they have at the moment is to strike the right balance between these two. What are the main difficulties to strike the right balance between innovation and protection?

- **Need to put more emphasis on innovation as a principle.** He made a first point saying that Europe has developed its prosperity over the last years thanks to progress and innovation but stressed that faith in progress has a bit diminished nowadays. There is a risk that we are turning too much on the precaution side[2]. According to Peters, it reflects where we are in the society: maybe we are too much cautious vis-à-vis innovation. It may undermine Europe's capacity and Europe as a whole to make progress and innovate.
- **We are in the middle of a huge transformation which is likely to be like a technical and industrial revolution**. In such revolution, you have to ensure 2 things[3]: (i) ensure that competition remains (because of the tendency of monopolization of the resource by the winners) which means how can we ensure that there is a vivid capacity of the industrial network to fight for competition (ii) ensure protection law for people: when you have losers and winners, you need to protect the society of the impact. Need to think our actions in this regard, both on how to ensure innovation and protection.
- **Date will be at the centre. Getting the data will mean getting the power**. In the next 20 years, there will be no sector that won't be impacted by the use of data and by the need of data. It will have a major

[1] Michal Boni referred to an article on the NYT on Facebook cooperation with its clients

[2] He referred to the speech of Jacques Attali in front of the French Senate where he stated that at the time when we invented automotive, if the precaution principle had existed, we would have forbidden it

[3] He made a parallel with what Bismarck or Roosevelt did 100 years ago when they had a new resource such oil

impact on value chain and the speed of the challenges is huge. Figures show that 2 years ago: 90% of data were not created. It is a deep transformation of the economic system. He stressed that things are changing in our real lives much more rapidly that our capacity in the political area. If everything goes well there is 5 to 7 years' time when we decide to do a regulation and we implement it. This mismatch becomes more and more important.

- **New issue such as data ownership to be addressed in the Strategy**. Need to make sure that the data element meets 3 requirements: (i) trust (otherwise there is no digital economy), (ii) innovation (need to make sure that business have access to data, huge source of innovation, competitiveness), (iii) public policy (data for protecting you but also data coming from a connected car could be useful to regulate traffic, should be useful for cities and public authorities). This triangle is not really set, so there is a need to prepare the ground for data to reach those 3 objectives.
- **The DSM strategy will be different than the previous initiatives in the area for 2 reasons**: (i) we want to have it focused (focus around 16 actions that we think we can deliver) (ii) we are ready to go on the direction of changes in many areas such as ecommerce, copyright, geoblocking, security, on data (ie: in ecommerce, we know where we need to focus and where we can add value). But there is a series of issues where we need for evidence (i.e. on platforms, we see the issues but we still need to gather more evidence). That will be a package with a series of issues where we think we can deliver in the 5 years' term and a series of issues where we need more evidence (i.e. on data: free circulation and cloud: need to make sure that people are trusting cloud, address standards issues)
- **Implementation is key and the DSM strategy will be a starting point**. There will be a big debate with huge stakes. The EU has something to say to the rest of the world. While we should remain open, we should also defend a high level of standards on data, on the way we can ensure fair competition between platforms and SMEs, on the way we see the European market progress on telecom. He strongly believes that there is some room of manoeuver, to make sure that there is more a European market rather than a fragmented and segmented market (for instance, on audio-visual services, there is a need to make the current system evolved).

To conclude, Eric Peters said that after the COMM delivers this ambitious message, it will be up to member states and to the European Parliament to take on what will be proposed and make sure that we can deliver on time, not waiting for another 5 years in a complete different world.

Benjamin André, CEO & Co-Founder, CozyCloud

First of all, Benjamin André introduced himself as an engineer and a startuper, not a lawyer. He presented his point of view regarding the context, stressing two main issues: privacy challenges and oligopolies. According to him, the source of the problem is linked to the centralization of data (centralization of the personal data, of the personal data governance and of the web), which explains the problems around privacy and oligopolies. In this regard, he highlighted two approaches:

- **The global approach**

Benjamin Andre argued that the market (meaning online services around personal data) needs fluidity. But today the **market is broken**, because big companies buy their competitors and aggregate the power around their products. He made a parallel with telecom industry and his own example: In France, in order to bring fluidity to the market, they try to help users to change of telecom operators. The regulator said that you phone number portability was the answer. According to him, we must find the same solution for online services. Fluidity means you should be able to change of services and in order to do so portability of data is key. If it stays as difficult as today to move from a service to another (your mails, contacts, photos, agenda…), you will never have fluidity and the market will remain broken whatever the tax rules or having 3 or 4 "Google" will not make any difference. He put the emphasis on **portability**.

Portability should be defined at a technical level. As a tech guy, he considers that "**Code is law"** and stressed the need to work on a code level. You must translate into law what portability means at the code level. You need to find some words in your law so that it influences the code. If you remain too high level in your human words, then your machine code will not been influence by your lawyers' words. Coming to the Data Protection Regulation, he pointed out that portability in the current text will have no effect because it lacks of code words. Because you need to influence the code and to do that, you need to talk about the code so that's why details and modalities matter.

If you say you can access the data: you need to **specify how to access the data**. He stressed that the format needs to be adapted to the content, so you need open and standard formats, and those formats must be machine readable. If a code cannot work on the content, the data disclosure doesn't work. It must be complete, exhaustive, free and delta mode. He stressed that those words are not so technical. Those words can influence the code, which is ruling the data. That is why in France they wrote a law proposal with Orange and the French Federation of Telcos, OVH (hosting company), FNIC and other start-ups to put those words into law. They gave it at the beginning of 2015 to the state secretary in charge of digitization in the perspective of her future digital bill. Maybe the government will build on that. They have a good understanding and they welcomed the approach. It is supported by small and big companies. He made a point on how can we influence Europe? He is not very optimistic, according to him it is obvious that a start-up cannot influence EU lawmakers but other big companies could spread the word and it will come to the text. According to him the **EU GDPR is not detailed enough. It will have no effect.**

- **The personal cloud approach , CozyCloud**

Then, he presented the personal cloud approach. He started by presenting the context of the web today struggling between privacy issues and oligopolies. He focused on the economic consequences of centralization of data. According to him, if you solve economic issues, you can also solve the ethical and privacy issues. What are the problems on the economic field?

- **Oligopolies are never good for consumers** because of the restriction of the market and of competition, innovation and for what can be offered by the market.
- **All pre-digital companies** - that started their business before digitalization**- are threatened by GAFA** (Google, Amazon, Facebook, Apple) and by the phenomena of "uberisation"

<u>How can pre-digital react to these disruptors? What are their options?</u>

- **Imitate:** they can try to create their own silo and try to aggregate as many data as they can. They can try to say "we can also to be a Google, Dropbox"
- **Find a new paradigm:** They would then need to innovate, find a field where they can be legitimate, have some advantages where the others cannot compete.

From his point of view, in the digital context, there are no options. Imitate is not possible because if you imitate, you will always be second or third since the winner takes all.

Pre-digital companies have to find a new paradigm. Benjamin Andre worked 10 years for a big company and said when he left he thought he will never be able to deal with big companies ever again. He noted that the context has changed in 3 years since every big company know they are not too big to fail. While it was impossible to talk to big companies 3 years ago, they now have start-ups incubators, funding … According to him many of those initiatives are just a trend but some of those are strongly linked to their strategy and the modification of their approaches and their future.

<u>Which new paradigm? Which requirements?</u>

- **Need to find a field in which they are not directly competing with GAFAs**
- **In this field, you should be able to deliver better services for the user**

<u>How can we achieve this paradigm?</u>

According to him, the way to achieve this paradigm is to give the data back to people "**data restitution to people"**. This model meets the 2 requirements: (i) it's the total opposite of GAFA's business model and (ii) it makes it possible to deliver better services (because you always need more data to deliver better services, more personalized and quicker services).

He stressed that the alternative approach to Google is to say: who is the most legitimate person to aggregate all the data of an individual? The answer is simple: the individual himself is able and legitimate to gather data. You **need to give him tools to manage his own data.** If you do so, you will create a space where you will have even more data than Google can have and then you can deliver better services than even Google can offer. It's the personal cloud approach where everybody could have its own cloud where he has all his data: mails, financial data coming from his bank and services, contact list, photos, health data… and then he will decide which service will work on his personal cloud.

He explained that CozyCloud is currently working with Orange, La Poste, and EDF (which are all pre-digital companies) and also has projects with some banks because they understand how they can benefit from this approach.

By definition, **a new paradigm must look utopian and surprising and must disrupt**. It took 50 years to democratize the PC so it may seem utopian to democratize the cloud. Regarding servers, things have not changed, you still need to be a high skilled person to interact with servers. The idea of the **personal cloud is to put a graphical interface on top of the server so you can administrate your own server** and to make it easier.

He made a quick demo to experience the personal cloud. He showed us how you can cross the data of your bank details and your mails or how the invoice of your telecom provider could cross the data with your contact list without having to hand over your data since the data remains on your server. The service is run in your cloud, under your control and your provider will never have access to your data but it can use the data to provide service.

He emphasized how this approach creates a lot of opportunities with some examples[4] that illustrate that on your personal cloud, you can have your energy provider working on Google's data but also access to data that even Google cannot access because on your personal cloud you have your bank account details, your agenda, history of many different objects of your Internet of Things and so on. The idea is to create a space on which pre-digital companies can deliver services that they could not deliver if they imitate today's paradigm of the cloud. The challenge is to gather data in the respect of privacy.

For the user, your personal cloud is the aggregator of your Internet of Things: it will allow your Samsung fridge to be connected to your IPhone. That would be the bridge between Samsung world and apple world). The user can make his own personal big data. **Your personal cloud is your digital home** and only people who are allowed can access to your data but without to be able to get the data. They don't have direct access. He stressed that the aim is to struggle against dispersion of data.

The ecosystem business model is around **coopetition.** He reminded us that in 2007, when IOS came out, phone manufacturers had 2 choices: they could build their own mobile operating system (Nokia, BB, Microsoft: failure today) or to federate behind a common architecture (Android: success).

In personal data, they have the same choice. You can remain in your vertical silo (finance/energy/entertainment). He stressed that the GAFA have a backbone of general services which allow them to go into vertical. The aim for personal data processors is to reach this cross-disciplinary approach. But to do so, they need to cooperate otherwise they will remain isolated. They cooperate in a way where every new service will bring added value to the other.

[4] For instance, your energy provider cannot access to your NEST thermostat data because of the General Conditions of Use set by NEST which state that you are not allowed to develop commercial access around your API. But if the user, the owner of the data, connects to the API and gathers the data, he can share the data with his energy provider.

To conclude, he said that with the personal cloud approach, data is physically and technically owned by the user. He called on the regulator to help to create the right conditions for fluidity of data so this kind of initiatives can happen. According to him, this could be a future of the web.

Discussion

- **Access to data and standardization**

Access to data happens to be conditioned by technologies and standards so which standards do we need in Europe? For instance, encryption is giving us tools for privacy and enabling data flow but it's also limiting access only to people who have the keys to do that. All the traffic is going to be encrypted and nobody will be able to see the traffic. It will be happening before you have your personal cloud so it's going to be useless if everything is encrypted.

MEP Boni stressed that there is a need for European interoperability and standards, which requires member states cooperation. Need for access and standardized access. He said that the problem of standardization should be deeper analysed, as well as the question of European interoperability, which are much needed to achieve a digital Europe.

Eric Peters stressed the need to ensure free flow of data and make sure that member states are not implementing measures that are preventing free flow of data. It also requires also having high standards: the idea that the Commission is trying to push is a complementary approach to the bottom-up approach of what comes out from actors in standardization. He said the COMM might try to have an up and down approach and identify what are the gaps, the priority needs for standardization. The approach the Commission would like to take is to see all the obstacles to free flow data and try to address all of them including the issues of standards.

- **GDPR as a burden for businesses?**

MEP Boni agreed that in some areas of the GDPR, MEPs should be much more cautious and should accept the position of small businesses. He acknowledged that small businesses will not be ready to fulfil all the obligations. This issue will be discussed during the trialogues. He stressed the need to adjust the GDPR to real businesses and make it implementable also for small business without burdens but with some obligations. He pointed out that the digital age creates complete new opportunities for personalization (such as the concept of citizens' energy managing). Besides these personalized services, it is also important for participation in democracy by using those new technological tools. As an institution, he said that the European Parliament is not ready yet for this idea of personalization but called on starting the discussion. He also defined the new model of definition of privacy, arguing that 30 years ago privacy was related to our body but now it's much broader. He stressed that people don't understand that, especially young people or parents when sending naked photos on YouTube. Saying that, he also addressed the problem of being forgotten if possible. In this regards, he said that regulators should cooperate with business and pointed out that this Regulation should not be seen as a burden but as an opportunity for better relation with clients and better personalized services.

- **Data ownership**

The audience addressed the issue of data ownership. Who owns the data? If you sell the ownership of your data: which protection do you have afterwards? Should data be open to everyone?

MEP Boni stressed that there are several kind of data and sometimes, for public services for instance, data should be open to everyone when it's not sensitive (public transports timetable). He then argued that there is a need to anonymize data when it's for an economic purpose.

On the issue of selling your data, he stated that the data of only one person is not really useful to business because they can't really build on that. However, he said that you should avoid selling your own data. From his point of view, the data protection rules should allow you to say "*I can agree for processing my data under*

rules but my data is my data". But, if you agree for something additional, you agree for something additional.

Benjamin André answered the question of ownership saying that defining what belongs to who is difficult. He made the parallel with the Intellectual Property Rights. He identified 3 uses for data: (i) selling (for commercial purposes); (ii) personal use (consumer's empowerment) and (iii) use by third parties. According to him, the most important is to organise the concept of ownership and we should focus on the two latter uses. Regarding the sharing of the data, he made the same distinction as MEP Boni saying that there is public data (not sensitive) and commercial data that you should protect.

- **Interoperability and portability**

Benjamin André made a clear point on this question stressing that interoperability and portability are two things different. These two concepts are not on the same level. According to him, lobbyists keep mixing those two things together to create confusion and be sure that we are not moving forward.
He explained that interoperability, from a technological point of view, is the "holy graal", it is too complex to get. He stressed that even within a same company, making all the software interoperable is a huge problem. For instance, to be able to extract your content from your Kindle (Amazon) and to be able to read it on another device: the only thing you need is portability for digital content. Portability is enough, you do not need interoperability.

- **Trust, security and privacy**

How difficult do you find for consumers to feel confidence in using the cloud services? Concern on the Issue of law enforcement to access to evidence because of new technologies: we might *protect also criminals via security (encryption), especially regarding child pornography: if a criminal stores all the pornographic material in his computer and because of encryption you can't have access. How to strike the balance? Regarding the personal cloud approach and the business model behind, privacy is not an absolute right; sometimes you need data for public interest or for research for example. Maybe it is a too individualistic approach. Don't you open the door to another oligopoly? At the end of the day: data will be stored in a cloud.*

To address these issues, **Benjamin André** stressed the crucial need to avoid building a fortress, arguing that this is the best way to fail. He said that the most important in the cloud approach is to detect and react to problems. For instance, CozyCloud's motto is "*you will stay because you can leave*". From his point of view, the best protection for the user is that the provider will lose his client if it doesn't fulfil the requirement; it is the best way to secure your data.
On collecting data in a public interest, he took the example on health and introduced the concept of peer to peer big data that is developing. It consists of a network where each nod can contribute to the computation. Everyone can contribute. It is a big data computation on a distributed map of nods.

On security and privacy, **MEP Boni** stressed that we need rules and under those rules we should process the data. Also, we should ensure law enforcement under strict rules (ex: with the PNR: after which time will we mask the data?). It is the same discussion on data anonymization for e-health and m-health. For instance, he said that for e-health, information about our health should remain between us and our doctor but on the other hand we need some knowledge about accidents, information on symptoms and so on for statistics and for preparing solutions. This is possible through anonymization: we can analyse it without personal identity.

From **Eric Peters'** point of view, the issue of security is more a technical problem rather than a policy issue. So far, the issue is linked to the capacity we have technically, when needed, to force the entrance of encryption.

- **Slow legislative process vs rapid technology changes**

MEP Boni acknowledged that technology definitely goes faster than regulation. That is why he thinks we should consider other options than regulation. According to him, we don't need strong regulation for all the digital single market. We need soft law (guidance, code of conduct, certificates related to standards) in

some area because it's easier to change when the technology is changing. He also stressed that the European Parliament's structure is completely unadjusted to the Digital Agenda: there are so many committees (IMCO, JURI, LIBE, ITRE and so on) working on it. So far, there is no possibility to organize a common special committee. He also stressed that this is impossible for the European Parliament to work faster so he calls for soft regulation (e.g. he referred to the success of introducing soft law with the example of the medical devices directive).

Eric Peters argued that the Commission doesn't want to ignore the democratic process which takes time. To increase the speed of the legislative process, the Commission focused on concrete and balanced proposals that the Commission thinks we can deliver by mid-2016. He then called on the co-legislators to tailor them to get an agreement.