The Impact of the General Data Protection Regulation on collaborative science in Europe and the European Cloud Initiative

Keynote Speech at ISC Intelligence in Science seminar, Brussels, 18 October 2016

Giovanni Buttarelli

Ladies and gentlemen,

First, let me thank you for your invitation. I am very glad to be here and to have the opportunity to share with you a few considerations on data protection and on the way it interacts with scientific research. The words “collaborative research” and “data-enabled science” have been used to describe such interaction. These are two expressions that effectively capture the added value of data, including personal data, for medical and scientific research.

In fact, if we look at how data sharing affects collaboration, we realise that data are, in the first place, a token for cooperation between researchers. I sometimes hear that research institutions are blamed for a certain reluctance - whether true or alleged - to share research data and experiences. In this respect, sharing data is, at the same time, an essential step for cooperation and a pledge of trust towards other researchers. In this sense, an infrastructure that allows data sharing facilitates exchange and reinforces mutual confidence and respect between research institutions.

From a different perspective, data are the fuel and catalyst of research - and here I refer to data-enable science. I am not a scientist, but it is obvious that in the twenty-first century science has

gained, or is equipped to gain, knowledge that seemed out of reach just a few decades ago. This depends, among other things, on big data, a combination of large sets of data and high computing capacity that allow us to draw scientific conclusions we could only imagine before.

Therefore, personal data lead to enhanced efficiency in scientific research, better understanding of diseases, new and more effective therapies, less people suffering, a longer life. The possibility of beneficial effects from personal data has not escaped to the Commission, which has recently launched a proposal for a European Cloud Initiative. Such European cloud has its cornerstone in the European Open Science Cloud, an infrastructure for the sharing of scientific data.

The idea behind the European Cloud Initiative is to make data available to the scientific community first and then, at a later stage, to the public administration and to business. This means that not only scientific research will benefit, but also public administration and business will experience efficiency gains. But then, we face one question: how does data protection fit the picture?

Well....we face a complex picture and I have, so far, presented to you only the brighter colours. The reality is that scientific research is a race against time towards a result and efficiency itself is a race. Racing is a difficult sport, dangerous and potentially lethal if you race your scooter against your friends on urban streets, less so if you do it on tracks with safety measures, an helmet and a car on four large wheels. To oversimplify, we are the ones providing the helmet and the other safety measures.

We supervise a safe and still value-adding use of personal data for legitimate purposes. We are enablers, as we ensure that efficiency is reached with less risk for personal data protection. We want the car to race faster, but we don´t want that car to have an accident, because all of us, citizens, are on that car.

Now, stepping out of my motoring metaphor, I would like to share with you the reflections that, at EDPS, we have carried out on cloud and data sharing and walk you through the personal data implications I believe we all should be aware of.

The most important issue, I believe, is accountability and the allocation of responsibility in a cloud environment. We will have someone using the cloud for enhancing scientific research and a cloud provider that knows nothing about research, but allows personal data (for example, my name, my blood parameters, my clinical history) to circulate faster and on a wider area. Who, between the two, should be responsible for complying with data protection rules?

In a simple world, the user of the data, the controller, would be responsible and the provider of the cloud, the processor, would just follow the controller´s directives. But in our world, the controller does not know anything about computing technology, about servers, security and encryption. These are subjects for the cloud provider. Therefore, we see that the traditional distinction between data controller and data processor, where the former gives directives which the latter follows is no longer adequate to the complexity of a cloud infrastructure.

That is why we need to re-think the user-provider relationship and consider them co-controllers, with shared responsibilities. In fact, if the cloud user is the one deciding for what purposes data should be used, it is the cloud provider to decide how the infrastructure should be designed and how data should travel on that infrastructure. That is sufficient, in our opinion as well as under the GDPR, to make it a co-controller.

Once we have determined who should be responsible for complying with data protection rules, we should understand which data protection rules should apply. In fact, the cloud infrastructure allows personal data to circulate as widely as possible, sometimes well beyond the borders of the European Union. Which rules then should protect citizens? The rules of the European Union? Those of the United States? Of China? Of any other country?

The most likely scenario, regulated by Article 3 of the GDPR, concerns the case where either the data user (the controller) or the cloud provider (the processor) have an establishment in the European Union. In such a case, provided that the processing of personal data occurs in the context of the activities of such establishment, the GDPR applies, regardless of whether the processing takes place in the Union or not.

There is then the case where neither the cloud user nor the cloud provider are established in the EU, but, in that case, I would tend to consider the rules on the transfer of personal data as the most effective tool to protect individuals.

There are, of course, other legal questions arising in connection with the sharing of personal data on the cloud, but my time with you is too valuable to be entirely devoted to legal intricacies. What I consider important is to deliver the fundamental messages that I hope will stay with you as you become active cloud users.

One of such messages is that the GDPR allows sufficient room for the development of science and scientific research, whether on cloud or not, through a number of well calibrated legislative exceptions.

For example, both the principle of purpose limitation and the related principle of storage limitation (*i.e.* data may be stored for no longer than necessary to achieve a certain purpose) are subject to derogation when a scientific purpose is present. We consider such derogation as an exception, to be interpreted and applied narrowly.

If we do not bind ourselves to a narrow interpretation of the exceptions provided for scientific research, we run the risk to become the whale hunters of this century, which, alleging scientific purposes, wander the digital seas to hunt personal data and then resell them for commercial use.

Another crucial challenge we need to face, in connection with the cloud, is that of data security.

Some of us, still in high school, may have come across Heisenberg´s *Principle of Uncertainty*: the more precisely we determine the position of a moving object, the more difficult it becomes to know its speed. I think this principle perfectly adapts to the cloud: the larger the circulation of personal data over the cloud infrastructure, the harder to monitor their position and use. The larger the access to personal data by multiple cloud users, the harder to ensure that all these users have in place adequate security safeguards, against unauthorised data access, theft or leak.

How should we address this problem? One answer is to correctly identify the entities which should be accountable for data processing, according to the principles I have described before. The other solution is *privacy-by-design*. This is a principle, now mandatory under the GDPR, that ensures that data protection safeguards are embedded in technology at the early stage of design and no longer treated as an afterthought. Of course, I need to add that *privacy-by-design* is often associated only with security measures, while it is a principle applicable to all rules of data protection in need of proper implementation.

I realize, to conclude, that many are the data protection implication of the cloud in relation to medical and scientific research and is impossible to exhaust all of them in a single intervention. For such a reason, I am taking this opportunity to mention that we are finalising an Opinion on cloud computing which will be published shortly. It is a substantial update of our Opinion of 2012, as four years are a geological era in technology. It includes some of the considerations I have just shared with you today and, more in general, elaborates on the application of the GDPR to the processing of personal data in a cloud environment. I hope you will find it an interesting and stimulating reading.