

# LEGAL & ETHICS

## The legal aspects of Big Data

Christian Laux

### Introduction Into The Legal Aspects of Big Data<sup>1</sup> (Part #1)

#### A. Notion and Definition

We start our analysis with a definition of "Big Data", in order to then determine what aspects are really relevant from a legal perspective.

##### 1. General

Big Data, as a term, describes the vast amount of data that are generated each day, worldwide, inside or outside of worldwide accessible networks, and that are generated on the basis of information taken from a variety of sources and stored in any format available, in order to then be processed and analyzed. When asking for the purposes of what Big Data aims at then the answer is data analytics, i.e. the exercise to distill meaning from data.



##### 2. Capturing Big Data through a number of "V's"

The interesting part in the term "Big Data" is that a quite structured approach for a definition is already widely in use, the definition relying on, depending on personal preference, on three or more "V's":

a) Volume, Variety, Velocity, Veracity, Validity, Volatility

- Volume: Big data implies enormous volumes of data. The reason for the wealth of data available in our networks is manifold. Nowadays, not people but machines create data (relying on user interaction or not). And the number of machines in use increases. Which, of course, will require more machines to read and process the data in use, and so on. Volume is a pointer to the „Big“-part in the term, but does not cause a lot of stir to anyone. It is just like that. Volume as such may be an organizational or computational challenge (the latter, not really), but, from a legal perspective, volume is not relevant.
- Variety: Variety refers to the fact that the data to be analyzed down the road has many origins, and the sources can be structured or unstructured. Data in PDFs, emails, photos, videos, and audio formats typically are unstructured. Data can, and by preference do, occur in structured format (csv, etc.). Variety as such is not a criteria that, per se, influences what Big Data means from a legal perspective. The variety must be considered, though, when we deal with the question how we can reuse the vast amounts of data we pull together, or find out there. Interestingly, the likelihood that data in unstructured formats is protected by copyright is much higher than copyright protection for data in structured format.<sup>2</sup>
- Velocity: Velocity means that data hits us at a high pace. The flow of data is not only fast, but also continuous, and researchers and businesses can increase benefits if they gain access to real-time data. Velocity is, per se, not a criteria that is legal in nature.
- Veracity: This aspect of the definition refers to the fact that not all that is being sent reaches the recipient free from noise, bias, or other shortcomings. Data being stored, mined, linked and then processed can be reused in a context that has nothing to do with the context in which it has been created. Data, per se, is neither true nor false, a claim that remains to be proven, at a later stage. But the veracity criteria is a pointer to the challenge that people need meaning, not (only) data. Big data refers to the exercise of extracting meaning from data, i.e. to data analytics as an entire discipline. We should say, data analytics is the very point of Big Data.

- **Validity:** Like veracity, validity asks for whether it is legitimate for a recipient to reuse data in a given context. The question is whether the data is correct and accurate for the intended use. Clearly valid data is key to making the right decisions. Legitimacy is obviously relevant for the legal analysis. We will keep this item to discuss it later in more depth<sup>3</sup>.
- **Volatility:** In this world of fast changes data yesterday's data may not be relevant to draw an analysis for today's challenges. Volatility, from a legal perspective is closely related to the challenges in Veracity and Validity.

## b) Nexus and Scale

To sum up, 'Big data' is all about scale and nexus which means to acknowledge that "datasets and databases are becoming so large and interconnected as to be virtually limitless"<sup>4</sup>. But we can, in the same time, draw the interim summary that the many V's of the definition in Big Data somewhat describe the phenomenon better than the legal implications.

That is why, we should add additional thoughts to determine what really matters, from a legal perspective.

## c) Small Data is Beautiful, too

One of the additional thoughts will help us to determine that "Big" is or should not be relevant. As Rufus Pollock, from the UK not for profit OKFN (Open Knowledge Foundation) points out, we should get rid of discussing only Big Data. In a blog post and an article with the Guardian, Pollock discusses why the current hype around big data actually misses the point<sup>5</sup>:

*"Meanwhile we risk overlooking the much more important story here, the real revolution, which is the mass democratisation of the means of access, storage and processing of data. This story isn't about large organisations running parallel software on tens of thousand of servers, but about more people than ever being able to collaborate effectively around a distributed ecosystem of information, an ecosystem of small data"*

While small data have a lot to tell us, there is still a parallelism in Small Data and Big Data.

It is the data analytics part that remains. The interesting part in Big Data as well as in Small Data is what we can do with data we have access to, i.e. the focus is on access and reuse.

## 3. Refining the Legal Meaning: What Really Matters

Parting from the above, we can hold that what really matters, from a legal perspective, is the following:

### a) Relevant Tests

The following steps are relevant when analyzing the legal implications of Big Data:

### Short bio

Dr. Christian Laux, LL.M., is Zurich based attorney for technology, advertising and e-commerce matters, combining both inside and outside counsel experience with a passion for technology.

In his daily practice Dr. Laux advises on all transactional aspects related to software, infrastructure and data distribution agreements, on matters of Intellectual Property, Telecommunications, Privacy, Corporate and Commercial Law. Litigation, in particular commercial litigation, domain name disputes and trademark litigation.

Dr. Laux has extended experience in data related issues, acts as data protection officers for companies, and is board member of Opendata.ch, an association dealing with data that are lawfully accessible to the public for the benefit of transparency, economical efficiency and more.

Dr. Laux completed his legal studies at universities in Zurich (Switzerland), Paris (France) and Stanford (CA, USA). His Ph.D. thesis discusses relevant issues about agreements covering copyrighted assets.

- Access: How can one lawfully access data?
- Use / Reuse: How can one lawfully use, process, analyze, and reuse data in new contexts?
- Deletion: How can one request deletion of data?
- Compliance: Is data one holds subject to restrictions that can be imposed on him by a third party or a regulator?
- Liability: How is one liable for the way how he analyzes, aggregates, uses or simply holds available data?
- Further aspects: There may be further aspects we will deal with separately.

## **b) Data (big or small)**

Big Data, as we have seen, is about data analytics. The task for data analytics is to extract meaning from data. Thus, the focus of Big Data, when looked at from a legal perspective, actually is (only) the "data" part of the term "Big Data". Whether the data are big or small is not actually relevant.

## **c) Meaning**

On the basis of data, then, meaning can be extracted. Extracting meaning can result in numerous implications, some of which can be tackled by the six prong test outlined hereinbefore. Personal data is about meaning, much more than about the very file that is put into a machine. Thus, the meaning layer of (Big) Data will need to be discussed in more depth, separately. But this would go beyond what we can discuss in this contribution. Let us save that conversation for later.

## **B. The legal landscape for data (including big data)**

### **1. A Kaleidoscope**

A colleague put it very well: "For all its growing significance in today's increasingly connected world, data remains funny stuff in legal terms."<sup>6</sup> The overall term to include data and meaning is "information". Interestingly, we live in an information age but have no uniform law to determine the legal implications of data, and its meaning. Rather, the law is a kaleidoscope of somewhat isolated rules that, in their entirety, make what can be called the "information law". There are different kinds of rights that may arise 'in relation to' data, rather than simply of rights 'in' data (copyright law rules, trademark law rules, data protection law rules, competition law rules, confidentiality rules, and contract law rules). This often causes some complexity when dealing with legal aspects of data, or related to data.

### **2. Contracts are Core**

The layered structure of rules that will be needed to assess the legal implications of data result in one instrument for all who are involved in data transactions, or who deal with data: Contracts. Contracts (if properly drafted) confer strong rights against the contracting party (the lawyer says: the party in privity), but not against everyone (unlike, for example, a copyright). Good news, then, are that contracts can be made enforceable internationally. And it is relevant to note that the situation abroad is comparable to the kaleidoscope approach we find for Switzerland.

## **C. Summary**

Big Data, when looked at from a legal perspective, is a somewhat misleading term, as one is tempted to focus on "Big" part rather than on the "Data" part of the term "Big Data". What we have outlined above results in the following recommendation, at least in the abstract: We suggest that a lawyer should take a step back from the "V"-driven approach, and rather adopt an approach focusing on the "kaleidoscope" (as referred to above) and contracts in order to properly assess the legal implications of Big Data.

## Notes

<sup>1</sup> This contribution is meant as a first article in a series of articles to discuss legal aspects of data and data analytics. Following this first overview, the legal landscape should be analyzed in more detail, then, specific fields of special regulation should be identified. Dealing data quality aspects can then be discussed, same as aspects of how one can be liable for the first step that is relevant for Big Data, the disclosure of data sets to the public.

<sup>2</sup> This statement is true for Switzerland, and a bit less true for countries in the European Union. The EU has adopted the concept of a sui generis protection for databases which is lacking in Switzerland.

<sup>3</sup> Liability for data will be discussed in a later issue.

<sup>4</sup> Big Data, Kemp Little LLP, January 2013, page 2, note 3.

<sup>5</sup> Rufus Pollock, Forget big data, small data is the real revolution, in: The Guardian, 25 April 2013, available at <http://www.theguardian.com/news/datablog/2013/apr/25/forget-big-data-small-data-revolution> (visited on 20 February 2014).

<sup>6</sup> Big Data, Kemp Little LLP, January 2013, page 7, note 9.

## SWISS ANALYTICS EVENT

### Web & Analytics

Sandro Saitta



The 3rd Swiss Analytics Event took place at Hotel Moevenpick Lausanne on January 30th. The topic, Web & Analytics, was covered through three talks. The event was sponsored by Webrepublic ([www.webrepublic.ch](http://www.webrepublic.ch)). Here is the detailed program:

- Introduction, Sandro Saitta, President of the Swiss Association for Analytics (5 min)
- Related organizations (5 min)
- *Using Web Analytics to optimize your Retargeting*, Jean-Marc Vandenabeele, Managing Director at HIPPARCOS (20 min)
- *Launch Hard or Go Home! Predicting the Success of Kickstarter Campaigns*, Vincent Etter, PhD student at EPFL (20 min)
- Break (10 min)
- Sponsor presentation (10 min)
- *Knowledge in Search Engines*, Maciej Kurant, Software Engineer at Google (20 min)
- Panel discussion about Web & Analytics with speakers (20 min) [*cancelled*]
- Networking Apéro



Pictures and most of the presentations are available at <http://event.swiss-analytics.ch>. The password for the presentations is available on our LinkedIn group ([www.swiss-analytics.ch](http://www.swiss-analytics.ch)). More information about the next event will soon be available on our LinkedIn group!

Want to be speaker or sponsor for a future event? Send an email to [info@swiss-analytics.ch](mailto:info@swiss-analytics.ch)