

Volume 10, Issue 2, August 2013

**9TH INTERNATIONAL CONFERENCE ON INTERNET, LAW AND
POLITICS (IDP 2013)
BIG DATA: CHALLENGES AND OPPORTUNITIES
REPORT**

Mariona Rosell Llorens *

DOI: 10.2966/scrip.100213.288



© Mariona Rosell Llorens 2013. This work is licensed under a [Creative Commons Licence](#). Please click on the link to read the terms and conditions.

* PhD Candidate, Faculty of Law, Universitat Pompeu Fabra Barcelona.

¹ Information about the analysis of the 15M movement in Spain is available online in the blog of DatAnalysis15m research group: <http://datanalysis15m.wordpress.com/>

1. Introduction – IDP Conference

The 9th International Conference on Internet, Law and Politics (IDP Conference) took place on 25 and 26 June 2013 in Barcelona. The IDP Conference is an academic event held annually in Barcelona which focuses on Internet and Information and Communication Technology (ICT) related topics from both legal and political points of view. The organisers of the event are: the Law and Political Sciences Department of the Open University of Catalonia (UOC); the Internet Interdisciplinary Institute (IN3 - UOC); and the Internet Derecho y Política Journal (IDP Journal). The UOC has been organising this event since 2005. The venue for the 2013 conference was the Cosmocaixa, the Science Museum owned by La Caixa Foundation.

Each year, the IDP Conference focuses on one broad topic. This year, the topic chosen was: “Big Data: challenges and opportunities”. In order to ensure a diversity of approaches and disciplines within the event, it is left to the potential speakers to define which specific aspects they wish to tackle. This encourages presentations from not only different legal disciplines, but also those from political and social science. The presentations were selected from those submitted in response to an open call for papers, although the organisers also invited speakers for specific panels and for keynote events. The conference brings together senior academics with international reputation, researchers and professionals from outside the academic world, and also early career scholars and PhD candidates.

As previously stated, the topic of this year’s conference was “Big Data” (BD) which as a term means large amounts of data (petabytes or hexabytes). However, taken as a phenomenon, BD can be understood to mean the storage and the usage of large amounts of data as a result of the continuing development of the ICTs and the expansive use of the Internet. The conference aimed to highlight the challenges, worries and opportunities that such a new way of dealing with data is giving rise to. The event was held over two days, and each day comprised of four panels and one keynote address. The legal approach towards ‘Big Data’ was divided into four topics: Intellectual Property, Regulation, Privacy, and Criminal Law. The political science approach was based on the use of BD for research, topics were divided into: Political Analysis and Social Movements.

2. Summary of the panels

As noted, the IDP Conference comprised of eight panels and two keynote addresses over the two day event. Due to the extensiveness and the diversity of the content discussed during the conference, just a brief summary of three of these panels is included in this report. Further information on the conference is available online at <http://edcp.uoc.edu/symposia/idp2013/> where the proceedings and conference papers are available to download.

2.1 Intellectual Property

The panel on intellectual property involved three speakers. Intellectual property concerns relate to BD in the sense that large amounts of copyrighted materials are available online and they are at disposal of the users, who can use them and remix them for other purposes.

Dr. Pedro Letai, lecturer of IE Law School, spoke about Copyright Regulation's policy online, meaning the political and legal design of intellectual property laws in the online environment. Nowadays, due to the vast amount of material available online, it is very difficult to provide effective copyright protection. Dr. Letai argued in favour of introducing a more flexible system of limits and exceptions in copyright regulation in order to promote innovation and ensure protection, this would in turn allow for a tailored solution for each case. He suggested that, from a law and economics standpoint, an open norm for what should be considered copyright infringement would be more useful, the courts ought to be able to consider and balance the infringement against the fair use for each case.

Marc Mimler, PhD candidate at Queen Mary Intellectual Property Research Institute (CCLS - QMU), explored the potential problems posed by the upcoming and developing technology of 3D printing for Intellectual Property Rights. 3D printers are capable of printing three-dimensional objects and reproducing already existing objects. Mimler analysed if patent laws in the UK and Germany are equipped to deal with such technology and whether the files used for 3D printing are liable for indirect infringement in the case of patented objects.

Irina Baraulic, PhD candidate at Vrije Universiteit Brussel and member of its Law, Science, Technology and Society research group, focused on the relationship between intellectual property and privacy. Baraulic explored the concept of "intellectual privacy", a concept that refers to the personal consumption and exploration of creative work located online assuming that it takes place within the individual's private sphere. The concept was developed in the US and Baraulic discusses how it could be used in the European context taking into account intellectual property and privacy regulation. This approach towards using copyrighted material relies in the idea of self-development and freedom of thought expressed through personal creations something worth to be protected under the privacy label.

2.2 Regulation

In the panel on Regulation there were three speakers. Professor Chris Marsden, University of Sussex Law School, (with a paper written together with Dr Ian Brown, senior research fellow at the Oxford Internet Institute) spoke about "prosumer law"; this is a regulatory model which is both market-based and citizen-oriented. Prosumer law is thought to prevent big internet corporations from over-controlling the online environment by ensuring interoperability among users and companies within the Internet. Prosumer law, as a model, aims to provide holistic regulatory solutions in what regards to information treatment in order to make sure that open commons (information which is not owned) does not become privatised by the major networking sites.

Humberto Carrasco, PhD Candidate at the School of Law, University of Edinburgh, spoke about regulation as a mechanism to enhance competition amongst companies in the telecommunications sector to try to generate the provision of better services. He compared the evolution of the US, UK and Chilean regulatory models for telecommunications. Carrasco defended an "emulated competition" notion in regulation – a model based not only on economic objectives, but also on welfare objectives. This model uses both sectoral regulation and competition laws.

Dr. Julián Valero, Lecturer of Administrative Law at the Universidad de Murcia, focused on the amount of data the government has (or can have) about its citizens. Dr. Valero warned about the dangers of promoting efficiency in administration through higher use of personal data to the detriment of citizens' legal rights (according to Spanish Public Law). He suggested that a better mechanism to provide the public with transparency with regard to government's activities (both towards citizens' data and in decision-making processes) is necessary to ensure efficiency and to preserve citizens' rights.

2.3 Criminal Law

The criminal law panel was composed of two invited speakers' contributions. Professor John Vervaele of Economic and Criminal European law at Utrecht Law School and College of Europe in Bruges, focused on two challenges within the criminal justice system entailed by BD and data-mining development: challenges to the rights of citizenry and the blurring of the boundaries of the criminal justice system as *ultima ratio*. Prof. Vervaele also noted that there is a change in the paradigm of criminal justice. It is now more focused on "security"; proactive behaviour by authorities to prevent crimes is encouraged over reactive behaviour.

The development of BD and data-mining techniques enable authorities to monitor citizens' lives if their behaviour is recorded in a database. Under the label of "security", infringement of citizens' privacy may occur. There is a shift in the criminal justice system within the Information Society context in the sense that surveillance has become an assumed tool of criminal investigation. BD and data-mining facilitate control and they entail a major change in what is thought the authorities should be doing to ensure security -that is demanding higher "preventive" control. As a consequence, criminal law influence is expanding among all areas of life, something that challenges its *ultima ratio* feature.

Dr. Ivan Salvadori, lecturer of Criminal Law and Criminal Computer Law at the Universitat de Barcelona, spoke about privacy protection in "The Cloud", and about prevention of, and sanctions for, illicit access to the information hosted in "The Cloud". Such issues are relevant since "The Cloud" is an information storage system not located in the actual computer or device of the user but in a remote server accessible through internet connection. International legal grounds for such privacy protection can be found in the Convention of Cybercrime of the Council of Europe of the 23rd November 2001.

Dr. Salvadori argued that according to current legislation (European, Spanish and Italian) such sanctions are possible provided that security barriers have been broken. Privacy protection in "The Cloud" is based on the Right to Informatics Privacy; this is a "new" and comprehensive right which comprises three aspects. It includes the content protected by personal information privacy (informational privacy); information under duty of secret of the data hoster (personal information and the communications' content); and other content hosted in "The Cloud" such as documents created, collected and edited by the user.

2.4 Other panels and contributors

The remaining panels of the conference were focused on privacy, political analysis and social movements. Privacy is one of the topics most challenged by the BD phenomenon, as a consequence ten speakers focused on one or another aspect concerning privacy issues, their presentations were divided in three panels.

In addition to legal concerns, BD as a methodological tool is also relevant for political and social sciences. The panels on political analysis and on social movements showed that BD is relevant for political behaviour understanding and also for social action analysis such as the 15M movement in Spain¹.

As noted above, further information on these panels, the speakers and their papers is available online in the IDP Conference official webpage.

3. Keynotes

3.1 Mireille Hildebrandt – *Slaves of Big Data. Are we?*

The first keynote speaker of the IDP Conference 2013 was Mireille Hildebrandt, Professor of Smart Environments, Data Protection and the Rule of Law at the Institute for Computing and Information Sciences (iCIS) at Radboud University Nijmegen. During her address she pointed at relevant philosophical concerns with regards to the BD phenomenon. She referred to literature that tackles the issue of Big Data, with special mention to Viktor Mayer-Schönberger and Kenneth Cukier's recent book: *Big Data*² (2013).

Prof. Hildebrandt spoke about the implications of the assumption $n=all$. Nowadays, thanks to the large amounts of data collected and stored online, and also due to the techniques of data-mining, there is a tendency to think that the population for any investigation or study equals "all of us" ($n=all$). As she noted, BD is a game changer because it changes our perception of what knowledge is, now it is seen as based on data instead of on explanatory relations between concepts and events; in a nutshell, with BD knowledge is much thought to be based in correlations, based on *what* happens, rather than *why* does it happen.

All this raises ethical and epistemological issues. It raises ethical issues such as monetisation of the personal data in addition to potential inequalities and asymmetries between corporations and individuals (and also between individuals). In addition, there are epistemological issues such as: "end of theory" –the end of classic methodology based on elaborating and testing theories on and about a limited population-; the lack of robust knowledge construction –knowledge stops to be conceived as stable, since it changes as the database used to generate it changes-; and, finally the epistemological issue due to the challenges posed by the Thomas'

¹ Information about the analysis of the 15M movement in Spain is available online in the blog of DatAnalysis15m research group: <http://datanalysis15m.wordpress.com/>. The analysis was undergone through processing large amounts of data (taken mainly from twitter) in order to find out the evolution, the nature and the characters of the movement, which occurred both within the internet and in the physical space.

²Mayer- Schönberger, V. and Cukier, K., *Big Data: A revolution that will transform how we live, work and think*. (London: John Murray Publishers, 2013).

Theorem. This sociological theorem states that something becomes real as it is thought real in its consequences, so rather than explaining the reality, reality is “created” according to the consequences taken from it, which in that case are the conclusions taken from massive data collection and data-mining processes.

Big Data implies a major change in science and business-making, indeed. Thanks to BD and data-mining, it is the actual behaviour (which can be monitored) that is relevant in designing businesses and/or governments’ action. As a consequence, there is no need for motivation in adopting one or another decision because justification derives from data predictions. So, despite the fact that BD facilitates better and more tailored solutions, Prof. Hildebrandt put forward two suggestions which may be useful in relation to the BD phenomena: firstly, there is a need for an effective right of data management; and, secondly, it might be desirable to reintroduce uncertainty in the era of BD.

To elaborate on these, the first suggestion was about the development of an effective right of data management, especially with regard to data qualified as behavioural and inferred (not voluntary provided by the individual and/or citizen). Prof. Hildebrandt claimed there should be higher awareness of which data is collected, how is it weighted, and for what purpose the data has been collected. More importantly, she maintained that people should engage in the right to access their data. Taking a Rawlsian standpoint, Hildebrandt suggested a distributive justice principle based on equity for this case. That is, provided that the level of welfare ought to be measured taking into account the level of welfare of the least favoured and; provided that everyone should have an equal portion of welfare, it would be fair that if someone increases the level of the community’s welfare, she has the right to have a bigger portion of such welfare. In the case of BD’s era, the one who significantly contributes to increasing the welfare is the citizen who provides the data, who, by now, is the less favoured because they are the ones with less control of the data being processed and less capable of autonomous decision-making.

The second important idea that Prof. Hildebrandt put forward was to question whether citizens’ free will might be affected due to the extensive use of BD and data-mining to design policies and other outcomes from governments or corporations. Her worries point at the potential alienation of the citizenry as governments’ and/or corporations’ have the ability to provide us with the “best” solutions according to our previously monitored behaviour, avoiding, in their turn, criticism or reactions against their decisions. The thing is, when promoting or enabling an auto-completing environment thanks to BD, expecting smooth compliance, there is a peril of alienation and over-control. Therefore, to prevent over-control, she suggests that the notion of transparency should include the idea of uncertainty –uncertainty when adopting a decision and uncertainty in citizenry reactions.

3.2 Duncan Watts—When does size matter? “Big data”, the Web and social science.

The second keynote of the IDP Conference 2013 was given by Duncan Watts, the principal researcher at Microsoft Research, founding member of the MSR-NYC Lab, former lecturer in Sociology at Columbia University (where he is currently a visiting lecturer), and visiting lecturer at Nuffield College, Oxford.

Instead of illuminating legal aspects, Watts' keynote focused on the potentialities that BD has for social research, however he also pointed to the fallacy in relation to prediction that the BD phenomenon is promoting within this type of research. In other words, on the one hand BD and data-mining techniques help to design and collect data in experiments and observation studies. On the other hand, it does not mean that BD facilitates accurate predictions, in fact, it shows that there is a high degree of unpredictability and a lack of causal explanations in observed behaviour and viral phenomena within the online context.

The first part of Watts' exposition showed there is a high degree of inequality and unpredictability in cultural markets (books or songs that become successful, for instance). BD techniques together with social research may help to rule out the assumption "x did better because x is better". With an experiment designed to explain the process of success (MusicLab³), he showed how data related with the "social influence" which a potential consumer receives (such as ratings or number of downloads), highly influences the subsequent success. In a nutshell, the bigger gets bigger. Therefore, we are faced with what Watts called the "Paradox of Social Influence". That is, despite the reality that individuals have more information upon which to base their decisions, choices taken collectively reveal much less information about individual preferences—they are more influenced by the previous social data rather than by independent variables such as, in this case, the names of the bands or the songs.

The second part of Watts' address focused on the degree of success predictability by using BD research. As he said, if we take the patterns of large cascades (viral success) and the patterns of small ones, we can see that they follow the same structure; therefore, online success cannot be predicted from its pattern of spread either because of the nodes. From his observation, Watts warns that marketers who traditionally looked at "who are the influencers" (assuming that some people achieved the degree of being influencers and others did not) are starting from a wrong assumption, because the degree of "success" of a cascade could be random in every instance.

His concluding remark was that BD management enables researchers to design accurate experiments, to collect and process data in order to test theories and/or to rule out assumptions—in other words, to balance theory with observation and experiments. However, it does not mean that they would be capable of fully predicting behaviour.

4. Conclusions from the rapporteur

The IDP Conference 2013 discussed the Big Data phenomenon from many different perspectives. It successfully highlighted the potentialities of the use of BD and data-mining techniques, and also served as a space for reflection on the challenges that BD entails for personal rights. The contributors to the IDP Conference all showed a special concern for the role of the citizen, both in being able to control personal information, and in being empowered to act freely, given the power that many internet corporations and governments have.

³ Information about Music Lab experimental study (papers and data) is available in one of its team members' webpage, Matthew Salganik (Princeton University): <http://www.princeton.edu/~mjs3/musiclab.shtml>.

In the legal field, the main conclusions which could be reached were the need to promote the notion of transparency and to design regulation which empowers the individual-citizen who is not only a consumer or a passive subject. Transparency has to be included (legally) in the treatment and storage of data, and in the procedures of public administrations and corporations. At the same time, it is necessary to ensure that individuals' rights are not being jeopardised, either by governments' actions or by the actions of the corporations. Indeed, some actual challenges are arising due to developing technologies, therefore, legal scholarship ought to be more concerned with *how* to protect data, *how* to preserve human rights, and *how* to enable interoperability from a sensible standpoint rather than being so much focused in defining *what* to protect. I would suggest doing so by accepting that the online context is not equal to the physical context and, therefore, the "problems" which arise might need different solutions.

For social and political research data-mining's potential is also crucial—especially with regard to the current social movements that develop in hybrid spaces (online-offline). The user-generated data ("entries", "shares", "likes" and others) represent citizens' actions in the online environment. When it comes to political science, these forms of actions, collective or individual, are relevant for researchers in checking the levels of legitimacy and/or acceptance accounted for by the political and legal system, and for designing more accurate policies corresponding to the demands of the population. BD, on the one hand, has methodological potentialities, because researchers have a rich and flourishing source of data for social and political studies. On the other hand, BD also has social potentialities, since it enables massive coordinated actions, which at the same time, represent singular personal expressions.

To conclude, it has to be said that since non-voluntary data is collected, stored and continuously mined, surveillance worries are especially relevant in what relates to self-development. Managing BD not only entails a technological development for better tailored solutions but also implies some deep questions about the actual freedom of choice and thought of the individual-citizen. In other words, the offer range destined to one individual is reduced to her potential interest according to the data previously monitored. Such procedure facilitates choice in the vast amount of options we currently have, however it is done by excluding the individual from the decision-making process and giving rise to some ethical and/or political problems.

Moreover, the fear that some particular inputs might be used to achieve desired reactions from the consumers and/or the citizens also exist, this poses the moral question of the existence of free will. It is this author's opinion that quite a few moral, legal and political questions should be tackled in relation to the BD phenomena, all in order to ensure that the ICT's potential abides by human rights and the rule of law.