UNIVERSITY
OF TRENTO - Italy
Faculty of Law
Department of Legal Sciences

lawtech

Trento Law and Technology Research Group

Student Paper n. 26

Big Data: Privacy and Intellectual Property in a Comparative Perspective

FEDERICO SARTORE

a cura di Roberto Caso e Paolo Guarda

ISBN: **978-88-8443-534-7**

KEYWORDS

Big Data – GDPR – Data Protection – Intellectual Property – Comparative

About the author

Federico Sartore (fgsartore@gmail.com) graduated in Law, *magna cum laude,* at the University of Trento, under the supervision of Prof. Roberto Caso and Dr. Paolo Guarda (December 2015).

The opinion stated in this paper and all possible errors are the Author's only.

ABSTRACT

Big Data is the fastest technology trend of the last few years. Its promises ranges from a philosophical revolution to a massive boost to business and innovation.
These great expectations come along with risks and fears about the dissolution of the traditional categories of privacy and anti-competitive effects on business. In particular, the dark side of Big Data concerns the incremental adverse effect on privacy, the notorious predictive analysis and its role as an effective barrier for the market.
The first stage of the legal analysis consists in an operative definition of Big Data, useful to build up a common background for further legal speculations.
Data deluge, the exponential growth of data produced on a daily basis in every field of knowledge, is considered the base for the existence of a Big Data world.
As a result, the practical applications of the data analysis involve healthcare, smart grids, mobile devices, traffic management, retail and payments. Moreover, the role played by open data initiatives around the world may strongly synergize with Big Data.
The main issues identified are studied through a comparative analysis of three different legal systems: US, Canada and EU.
Notably, the origins of privacy in the US are considered to sketch the line toward the US policy is moving. On the other hand, the current draft of the General Data Protection Regulation on EU level is completely changing the landscape of data protection.
Finally, the European influence is clearly perceivable on the Canadian legislation.
Although the level of protection granted slightly differ, it is still possible to identify the common consequences of the rise of Big Data on the legal categories.
In particular, the fall and redefinition of the concept of PII, the question whether the binomial anonymization/re-identification may still exist, data minimization and individual control. The attempt of this paper is to provide a multi-layered solution given to the so-called Big Data conundrum.
Consequently, the single layers are represented by: proactive privacy protection methods, self regulation and transparency, a model of due process applicable to data processing.
The second part of this paper is dedicated to answer a challenging question: whether or not IP traditional categories are suited to work with Big Data practices.
This section of the work focuses on the different practices used in the market before summing up the common traits. In this way, pros and cons of the application of the traditional IP legal constructs are considered having regard of a general category of Big Data practice.
Eventually, the lack in the current legal landscape of an IP construct able to meet the needs of the industry suggests to imagine the main characteristics of a new dataright.

4.3 A third way to disclosure

INTERVIEW WITH DOCTOR ANN CAVOUKIAN

Introduction

The 12th of December 1938 Enrico Fermi, in his Nobel Lecture *Artificial radioactivity produced by neutron bombardment,* summarized the results of his scientific research about artificial radioactivity induced by neutrons; in particular, he mentioned the artificial birth of two new elements of the periodic table: Ausonio and Esperio. Ten days after the Nobel Lecture, two German physicists, Otto Hahn and Fritz Strassmann, denied the discovery of these "transuranic" elements performed by Fermi and his team[1].

This anecdote is to recall the instinctive process, which binds the birth of an entity with the rise of its name and its fall/modification with its death/evolution. Beyond the age-old problem of arguing which came first the chicken or the egg, the time and circumstances of the first use of a name can provide useful information and insights on the phenomenon.

Tracing back the origins of the Big Data phenomenon starting from its intriguingly Orwellian name (especially if capitalized), one has to search for both academic/non academic and published/unpublished material. The term probably originated in lunchtime activity at Silicon Graphics Inc. (SGI) around the mid-90's with the prominent contribution of John Mashey[2], the first significant academic references can be addressed to Weiss and Indurkhya (1998) [3] in computer science and to Diebold (2000) [4] in econometrics/statistics.

Moving from the signifier to the signified, one may encounter many, various and almost impairing troubles struggling to give a univocal and unambiguous definition of Big Data.

---

[1] M.Leone and N. Robotti. "Enrico Fermi E La Presunta Scoperta Dei Transuranici." *Atti Del XXIII Congresso Nazionale Di Storia Della Fisica E Dell'astronomia* (2003). Available at http://www.brera.unimi.it/sisfa/atti/2003/231-244LeoneBari.pdf

[2] F.X. Diebold, *A Personal Perspective on the Origin(s) and Development of 'Big Data': Phenomenon, the Term, and the Discipline* (2013). Available at http://www.ssc.upenn.edu/~fdiebold/papers/paper112/Diebold_Big_Data.pdf

[3] S.M. Weiss and N. Indurkhya, Predictive Data Mining: A Practical Guide (The Morgan Kaufmann Series in Data Management Systems) (Morgan Kaufmann Publishers In, 1st ed. 1997).

[4] F.X. Diebold, *Advances in Economics and Econometrics: 'Big Data' Dynamic Factor Models for Macroeconomic Measurement and Forecasting: A Discussion of the Papers by Lucrezia Reichlin and by Mark W. Watson* (Publisher:Cambridge University Press 2003).

The initial idea lying above the name was that the volume of data had grown so large that the quantity being examined no longer fit the memory of the computer used to perform the task, so the engineers were forced to rethink new data crunching tools (Google's map Reduce and his open source equivalent Hadoop).

One way to tackle the issue today is to think is that big data refers to a scalar jump, in other words it refers to things that can be done at a large scale but not at a smaller one. This partial and seemly plain definition is about the what, but is the how that is generating new business models, innovation and a huge and far to be solved debate. It's the second death of God[5], the end of the scientific theory for some analysts[6] or simply the end of the theorist for others[7], the final stage of the epistemological process, which finds its roots in the Heisenberg uncertainty principle corollary of living in world ruled by probability rather than causality.

Following this route, we will enter a world of constant, pervasive but beneficial data driven predictions lead by the autocratic binomial data-algorithm, a world of correlation. What would it imply, for instance, a medical doctor who cannot justify the reasons behind a medical intervention because based on a big data high degree of correlation rather than on the time-honored cause-effect principle? In a correlation-driven legal system an individual could easily suffer imprisonment in advance for a crime he is predicted to commit.

Beyond these dystopian considerations, what does Big Data mean today? What typology of business models lies behind this name? What provisions are needed for a legal system, which aims to regulate the phenomenon boosting innovation and protecting competition on the market? How will the "old" concept of privacy – depicted as a useless *maginot line* – fit the attempts of the policymakers of defending their citizens and their free will against the dictatorship of data? What kind of legal tools will provide stability to legal systems that are, by definition, elephantine in keeping the pace of such a tumultuous mixture of innovation and individual behavior? Are the IP constructs well suited for "the change of skin" they require to face this brand new challenge?

---

[5] V. Mayer-Schonberger and K. Cukier, Big Data: A Revolution That Will Transform How We Live, Work, and Think p.8 (Eamon Dolan/Houghton Mifflin Harcourt 2013).

[6] Wired and C. Anderson, The End of Theory: The Data Deluge Makes the Scientific Method Obsolete (WIRED Jun. 23, 2008), http://archive.wired.com/science/discoveries/magazine/16-07/pb_theory.

[7] I. Steadman, Big data, language and the death of the theorist (Wired UK) (Wired UK Jan. 25, 2013), http://www.wired.co.uk/news/archive/2013-01/25/big-data-end-of-theory.

In general, the age of Big Data will call for new rules and regulations to shield the individual and his rights. The creation and adaptation of legal principles and their specification through specific provisions is a process that is taking different shapes on the two sides of the Atlantic.

A consistent portion of the *Digital Agenda for Europe*[8] refers to data in its bilateral meaning: as indispensable assets for a flourishing economy and as endangered immaterial extensions of the European citizen, which need protection.

This protection has to be uniform to be effective. In this very moment[9], the EU institutions are focused on the draft of the new Data Protection Regulation that, being directly applicable with no need for a domestic transposition of the single member state, will consistently influence the data collection and reuse landscape from the day-one of its effectiveness. On the side of incentives for a growing economy, an important initiative taken at EU level is the PSI (on reuse of Public Sector Information) directive, representing the legislative option of the open data policy pursued by the Union; the policy is made complete by a number of non-legislative measures to support the opening up of the public sector information.[10]

Meanwhile, in the US the relevance given to the topic can be valued considering the policy of investment ($200 millions announced in 2012) in Research and Development on Big Data promoted by the White House.[11] in 2014 a 90-day study on Big Data was issued on request of the President depicting a whole world of opportunities to discover.[12]

---

[8] EU COM(2010) 245 final. Available at

http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2010:0245:FIN:en:HTML

[9] The legislative process entered its final stage with the so-called "trilogues" between the EU Parliament, the EU Commission and Council of the EU to agree on the final text of the regulation. The trilogue stage should end in December 2015 and the final text should be adopted in the first months of the 2016. This document is available at:

http://www.europarl.europa.eu/news/en/news-room/content/20130502BKG07917/html/QA-on-EU-data-protection-reform

[10] Directive 2003/98/EC consolidated by Directive 2013/37/EU.

[11] T. Kalil, *Big Data Is a Big Deal*, The White House (2012), available at https://www.whitehouse.gov/blog/2012/03/29/big-data-big-deal.

[12] J. Podesta and P. Pritzker, Big Data: Seizing opportunities, preserving values (May 1, 2014), available at https://www.whitehouse.gov/sites/default/files/docs/big_data_privacy_report_may_1_2014.pdf.

This thesis will try to give an answer to the above-mentioned questions, digging inside this protean phenomenon with a shifting definition and tiny boundaries that nowadays represents one of the fastest-spread trends for both the industry and the academy. An uneasy task that will be performed moving constantly back and forth from the world to the legal system and from the legal system to the implications and consequences it generates. The approach chosen to tackle a matter so deeply related to the existence of the information society will require scaling and considering the parallel efforts of EU and North American legal systems to cover with law the Big Data Golem in order to harness big data and not be crushed by it.

The structure of this work is made up of three chapters: the first will attempt to describe a multi-faceted phenomenon with the goal of delimiting the scope of Big Data in order to isolate the legal consequences and safeguards arising from it. This limitation task has been undertaken starting from the current definition of what Big Data is today; afterwards the focus has been moved to the potentiality of the technological phenomenon: the value of the data, the synergy with the diffusion of open data policy approaches and the risk and threats arising from it.
The second chapter represents the main body of this thesis, briefly investigating the past and new elements of three different legal systems in order to move to the study of the shifting of the classic privacy categories needed to counter the new challenges posed by the Big Data analysis. Mainly, the shift in the definition of PII, the new conception of the powers of the anonymization/re-identification dichotomy and the rising of technological approaches to privacy issues.
The third and last chapter will deal with the relationship between Big Data technology related practices and intellectual property rights, trying to sketch the basis for a further analysis about how to foster innovation in a field where the high pace and recycle of structures is leaving the "old" IP rights out of the playfield.
Finally, in the genesis of this work the research undertaken at McGill University in Montrèal has played a fundamental role giving me the chance of working side by side with legal students and scholars who are living, studying and working in a different and complementary legal system. Thus, the shape and structure of this thesis has been influenced by the Anglo-Saxon style of draft and research.

I. New engines for new oil – the present of Big Data

1. Singing in the data deluge

In Latin the word *data* means *given,* in the sense of a fact. The current definition is nothing more than an extension of the Latin one, nothing more than a (given) description of an entity that allows it to be recorded, detached from the material landscape and so re-organized and analyzed.

What lies beneath the surface of the explosion of Big Data is the modern expression, made possible through the development of the technological horizon, of a constant quest of humankind: measuring, recording and analyzing the world. This will and task of quantifying the world has been defined *datification*, logical premise for any data analysis scenario[13].

An idea of what distant shores this tension to ubiquitous measurement can reach can be given by the work of Shigeomi Koshimizu and his team of engineers at Japan's Avdanced Institute of Industrial Technology in Tokyo[14]. Koshimizu converted backsides into "smart backsides" filling them with sensors able to measure the pressure of the body in 360 different spots, processing them into a digital code that is unique for individuals. During a test the system recognized the individual from the posture with 98 percent accuracy; the practical applications of similar technologies could be really impressive.

Moreover, in this and similar cases there is something more: the data was born analog and become digital to receive a useful meaning for the analyst.

A common belief tends to put in equivalence, or at least, to join as two trees with common roots, the datification and digitalization phenomena; nonetheless their contemporary evolution is a vital requirement for a big data world, nowadays we assist to a shift of focus inside the IT revolution and precisely from the T to the I.

This mixture of new devices made available by the new technologies and old desire for measurement paved the way for a process, which brought quality out of quantity and gathered quantity through the evolution of computational power and storage availability of the machines. The outcome is a flow of data that outpaced the evolution of adequate

---

[13] See note 5 pag.78.

[14] N. Owano, Engineers unleash car-seat identifier that reads your rear end (2011), available at http://phys.org/news/2011-12-unleash-car-seat-rear.html.

infrastructures and tool to manage a data driven world[15]. It is the so-called *data deluge*, a redundancy crisis.

The data deluge finds in the lines above its historical and technological justification; however, its huge relevance can be better achieved and explained looking for some examples of what it means today.

In the first two weeks from his creation in 2000, the Sloan Digital Sky Survey collected more data than in the rest of the history of astronomy, after 10 years its archive contained 140 terabytes of information; and there is more, his successor, the Large Synoptic Survey Telescope that will start to operate in 2016 will gather the same amount of information every 5 days[16].

In finance, about seven billion shares are exchanged every day on US equity market; computer algorithms based on mathematical models trade two-thirds of it[17]. Internet companies are a perfect target to analyze in order to consider the size of the advent of the data deluge: in 2013 Google used to process a volume of 24 petabytes per day, thousands of time the quantity of information stored on the shelves of U.S. library of Congress[18].

Concluding this factual digression, the personal dimension of the digital deluge we are experiencing today is the equivalent of giving every person on Earth 320 times the information is estimated to have been stored in the Library of Alexandria (representing in the third century b.C. the sum of all knowledge in the world)[19].

A good way to tackle this gigantic increase is to ask the meaning of this deluge of information. It is the same as physics or nanotechnology: size matters. In other words, reading the words of Oscar Wilde as an unaware prophecy: "It is a very sad thing that nowadays there is so little useless information". Using a colorful metaphor he food at the

---

[15] *Leadership Under Challenge: Information Technology R&D in a Competitive World An Assessment of the Federal Networking and Information Technology R&D Program*(President's Council of Advisors on Science and Technology 2007). Available at https://www.nsf.gov/geo/geo-data-policies/pcast-nit-final.pdf

[16] J. Lewi and K. Cukier, *Data, Data Everywhere*, 3 (The Economist 2010). Available at http://www.economist.com/node/15557443

[17] R. Nazareth and J. Leite, *Stock Trading in U.S. Falls to Lowest Level since 2008*, Bloomberg (2012). Available at http://www.bloomberg.com/news/articles/2012-08-13/stock-trading-in-u-s-hits-lowest-level-since-2008-as-vix-falls

[18] T. Davenport et al., *How Big Data is Different*, Sloan Review (2012). P.43-46. Available at http://sloanreview.mit.edu/article/how-big-data-is-different/

[19] See note 5, p. 9.

researchers banquet is far beyond the capacity of their stomachs, however it is interesting to analyze every single course to get how it got to the table.

2. Defining Big Data

Trying to accomplish the attempt of defining this two-words phenomenon, which bears in itself the seedlings of an epochal revolution, it might be useful to collect some of the most inspired definitions.

Although no set definition exists[20] for big data (and for small data too) it is still possible to trace back the history of this shorthand collecting thoughtful insights from its evolution through the years.

The first definition of big data comes from a 2001 report of the META Group (now Gartner), the challenges and opportunities related to the data growth were described as being three-dimensional, increasing volume (amount of data), velocity (speed of data in and out), and variety (range of data types and sources)[21]. This 3Vs model became the standard for the industry referring to big data.

Later on in 2012, Garner updated its definition describing big data as follows: "Big data is high volume, high velocity, and/or high variety information assets that require new forms of processing to enable enhanced decision making, insight discovery and process optimization"[22].

Sometimes, a new V "Veracity" is added by some organizations to describe it.

The academic world, on the other hand, tends to be more verbose and a good example of these academic definitions might be the following: "Big Data is shorthand for the combination of a technology and a process. The technology is a configuration of information-processing hardware capable of sifting, sorting, and interrogating vast quantities of data in very short times. The process involves mining the data for patterns, distilling the patterns into predictive analytics, and applying the analytics to new data.

---

[20] J. Berman, *Principles of Big Data Preparing, Sharing, and Analyzing Complex Information*, Amsterdam: Elsevier (Morgan Kaufmann 2013).

[21] D. Laney, *3D Data Management: Controlling Data Volume, Velocity and Variety Meta Group Report* (2001) available at http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf

[22] M. Beyer, *Gartner Says Solving 'Big Data' Challenge Involves More Than Just Managing Volumes of Data* (Gartner 2011) available at http://www.gartner.com/newsroom/id/1731916

Together, the technology and the process comprise a technique for converting data flows into a particular, highly data-intensive type of knowledge"[23].

The idea behind this historical sequence and difference in background of the definitions above is to tackle the hard definitional task from many different points of view in order to depict a multidimensional portrait of the phenomenon.

2.1 Words as data

We described datification and digitalization as two distinct but connected phenomena, one of the field in which they unleashed their potential is the realm of the written words. The trailblazer is once again Google with its *Google Books* initiative in 2004; the goal was to take every page of every book and to make it digital in order to give the access to anyone with an Internet connection.

Nothing new under the sun (ignoring the scale of the project that employed digital scanning machines able to turn the pages making the task feasible), project Gutenberg, born back in the 70's, had the same core, however, the mission was "just" making the books available for people to read[24].

Google took a step further datifing the books once available in digital format; they implemented software to access them in a machine-readable format[25]. Datification made the text indexable and thus searchable permitting in this way to analyze with algorithms the dataset of human culture.

It is curious to notice that the possibility of pursuing this kind of search gave the sparkle of a new academic discipline: Culturomics (computational lexicology that tries to understand the human behavior through the words).

From the legal point of view issues may arise and arose for what concerns the re-use of information protected by copyright. In two famous cases the judges held that Google met all the requirements in order to be shielded under the *fair use* doctrine[26].

---

[23] J. Cohen, *What Privacy Is For,* 106 Harv. L. Rev 1904*, p.1920-21* (2013).

[24] J. Thomas, *Project Gutenberg Digital Library Seeks To Spur Literacy*, U.S. Department of State, Bureau of International Information Programs (2007). Available at

http://iipdigital.usembassy.gov/st/english/article/2007/07/200707201511311cjsamoht0.6146356.html#axzz3let5qx1Y

[25] J. Michel et al., *Quantitative Analysis of Culture Using Millions of Digitized Books*, Science p. 176-182. (2011).

[26] *Authors Guild, Inc v Google, Inc*, (2011) 770 F.Supp.2d 666 (SDNY) and *Authors Guild, Inc. v. HathiTrust,* (2014) 755 F.3d 87 (2d Cir*).*

2.2 Location as data

The important year to remember is 1978, in which the first of 24 satellites that make up the Global Positioning System (GPS) was launched. Its evolution is at the basis of the current possibility of retrieve the position of everything in the world. Today, an increasingly high number of applications for smartphones gather location information from the individuals regardless of any form of location-related feature of the application itself. The collection of locations is becoming a highly valuable activity, in particular whether the scale of information tends to ascend. A number of flourishing companies built up their business on this information (for instance AirSage created a real-time traffic plan based of 15 billion of mobile users records). Likewise, even traditional business models have not remained indifferent to the potential of geo-positioning.

The insurance system, for example, might suffer an actual revolution, at the moment in the U.S. and England it is possible to subscribe an insurance policy based on where and when the drivers actually drive[27]. The shift from the insurance contract as a way to pooling the risk to a contract tailored on the individual action could have deep consequences on the legal structures of these agreements.

2.3 Interactions as data

Although the leading position of social media companies in the world of big data is not part of the breaking news, however, the potential of the analysis of the data outflowing from all the daily interactions of the hundreds of millions of users of Facebook, Twitter, Linkedin, is a definitely a fundamental asset for the different firms operating in the market of the re-use and analysis of information.

Some hedge funds in London and in California started to analyze the flow of tweets, crunching the data to gather precious correlations between the usage of the social platform and the individual behavior of the human being or company posting a tweet[28]. Actually a case of tweet-driven prediction is rather famous: Bernardo Huberman, one of the fathers of the social network analysis, developed a model able to predict the success or

---

[27] See note 16.

[28] K. Cukier, *The Mood of the Market*, The Economist (2012). Available at http://www.economist.com/blogs/graphicdetail/2012/06/tracking-social-media

failure of an Hollywood movie at the box office taking into consideration the rate of the tweet posted[29].

The model turned out to be much more reliable than the older and commonly used predictors.

Beside the raw data expressed in a 140-character message metadata also count: users language, their geo-location, the number and identities of people they followed or they were followed by; a Science report claimed that mood has been finally datafied[30]

[29] A. Sitaram and B. Huberman, *Predicting the Future with Social Media*, Proceedings of the 2010 Ieee/wic/acm International Conference on Web Intelligence (2010). available at http://www.hpl.hp.com/research/scl/papers/socialmedia/socialmedia.pdf

[30] S. Golder and M. Macy, *Diurnal and Seasonal Mood Vary with Work, Sleep, and Daylength Across Diverse Cultures*, 333 Science (2011). Available at http://www3.ntu.edu.sg/home/linqiu/teaching/psychoinformatics/Diurnal%20and%20Seasonal%20Mood%20Vary%20Across%20Diverse%20Cultures.pdf

3. The value of Data

The main shift the value of information takes in a big data world is strictly related to the possibility of making a profitable use of it. Although the information has always had an intrinsic value, it was strictly related to the core business or at least to a relatively close number of specific sectors or narrow categories such as personal information or intellectual property.

The crucial difference of the current approach of the economic operators is that every single piece of information, even the most seemingly useless, is valuable in and of itself. It is a generally renowned fact that information, for its own nature, is what economists call a "non-rivalrous" good. The point is that the possibility of re-use was conceived as related to the same initial purpose; the difference with big data is that the first use of a dataset does not imply a full usage of the "asset". Furthermore, and consequently, the difficulty in estimating the value of a collection of data is the other side of the success of those start-ups that will be able to extract value from a "pit of wonders".

No rigid business models will survive the market of re-use of information for the only and simple reason that new perspective of the same object may result in a new factor of production to exploit. The main legal issue related to re-use of information from an economic point of view is to enforce legislation able to guarantee an equal access to the market of information to all the economic actors, prohibiting all those acts that express an abuse of dominant position.

3.1 The re-use of Data

One of the brightest examples of re-use of information is a business model that could be defined as an offline search engine; the idea was of the consultants at McKinsey & Co[31] who realized that all the data produced by an anonymous delivering company could receive a valuable implementation once aggregated. The result was a forecasting machine for business and international exchange of goods all over the globe.

Moreover, other economic operators with a good position in the information value chain can amass huge quantities of information they do not really need to run their business. For these companies, all the data they gather has only narrow technical uses.

[31] B. Brown et al., *Are You Ready for the Era of "Big Data?*, McKinsey Quarterly (2011). available at http://www.mckinsey.com/insights/strategy/are_you_ready_for_the_era_of_big_data

Mobile Operators for instance, and Telefonica first of all[32], discovered the potential of knowing where, when and what kind of strength had the signal during the telephone conversation of their clients.

3.2 The value of data exhaust

The re-use of data can sometimes take a different, hidden form, every single time a human being interacts with a machine for a purpose (s)he implements a number of actions that can be recorded and used for commercial and non commercial goals. This is essentially what the expression *data exhaust* means: it is the byproduct of the interaction of the people with the world, which surrounds them. For its online counterpart it describes every action they take: how long they stare at the screen before clicking, where do they move the cursor, how often they retype a word, etc.

All this apparently useless information has a value and a price. The other side of this huge competitive advantage for those companies, which can collect the data exhaust and have the insights needed to make them profitable is its existence as powerful barrier to enter against rival. It is a powerful know-how based on the customer experience and it takes its effectiveness from the scale it operates; that is why it represents a big data related issue and why new comers will not be able to offer the same quality of service.

Simply because the improvement of your service needs data and data are provided by clients, clients who will not choose the new service until the operator will not offer a product at least equivalent to those yet on the market.

3.3 Valuing a priceless entity

The difficulty of the task of valuing and immaterial asset is something taken for granted, expressing in monetary value the potential of the information is similar of what happened with the pricing of financial derivatives before the development of the Black-Scholes equation in the 70's or the troubles in pricing patents, field where the market is made up by auctions, litigations, licensing and private sales.

The traditional approach to the pricing of intangible assets tends to emphasize different elements that, albeit not being part of the formal financial accounting system, are

---

[32] BBC, *Telefonica Hopes 'Big Data' Arm Will Revive Fortunes,* BBC Technology, Oct. 9, 2012, available at http://www.bbc.com/news/technology-19882647.

nonetheless taken into consideration performing the task. This set of elements includes brand, talent and strategy.

Towards data there is no correct way to tackle the issue, different and varies are the strategies applied, their weakness in depicting the value of something potential was widely shown in occasion of the new beginning of Facebook as a public company in 2012. The case is paradigmatic: before the beginning of the IPO (initial public offering) the estimated value of a Facebook's share was 38$ for a total capitalization of 104$ billion with a divergence from the value of Facebook, according to the classic accounting standards (6,3$ billion), of almost 100$ billion[33].

One way to fill with rationality the huge gap (not only the Facebook one) between a company's "book value" (the value of all its assets according to its balance sheet) and its own "market value" (the price at which an asset would trade in a competitive auction setting), is to start looking at the different strategies data holders use to apply in order to extract marginal value[34]. In this way markets and investors will try to price these assets giving the possibility to the legislator to implement, once that accounting quandaries and liability concerns will be mitigated, new accounting rules to let the value of data show up on corporate balance sheets, lawfully becoming a new class of assets.

The first possibility in doing so is to look at the different policies data controllers and data holders apply in order to generate value, for example the most typical use is the possibility for the firm to consume itself the asset in its productive process. However, considering the importance we gave to the idea of data as a bottomless mine, it is unlikely that a company is capable of mining all the gold; to avoid an inefficient distribution of the factors of production the classical alternative is the license.

Licensing data could be not so difficult considering the past licensing strategies applied, for example, in intellectual property deals in biotechnology, where licensors can demand royalties on all the inventions derived from their first licensed technology.

In this case in fact every single party of the deal has an incentive to maximize the revenue related to the data reuse activity, however, considering the fact that we are dealing with non-rival goods, and the fact that is unlikely for a single licensee to exploit the full potential of a x dataset, an exclusive license could represent a good option only if the exclusivity clause was related to the specific way of reuse.

---

[33] D. Laney, *To Facebook: You're Worth $80.95* (WSJ 2012), available at http://blogs.wsj.com/cio/2012/05/03/to-facebook-youre-worth-80-95/.

[34] R. Kaplan and D. Norton, *Strategy Maps: Converting Intangible Assets into Tangible Outcomes.*, Harvard Business Review Press (2004).

In other words, in order to allow the data operators to hedge their bets when dealing with this new *data market*, "data promiscuity" might become the norm, and a winning one.

Meanwhile, different companies had tried to enter the market of data getting free datasets, re arranging them in a easier way and reselling the: one of these is DataMarket, founded in Iceland in 2008 (or Windows Azure Marketplace, a Microsoft branch dedicated to a high-quality data offer).

4. New models for business and research – seizing the initiative

It is easy to understand from the lines above that Big Data means big industry, and different researches show it; the companies that use forms of data-driven decision making enjoy a boost in productivity around 5%[35]. Moreover, it is crucial to underline that it is not just a treasure chest ready to be opened by companies' executives, governments are enjoying the beneficial effect of big data too.

It enhances the public sector administration and can assist global organizations in crunching data to improve their strategic planning.

This section explores the practical benefits for business and research with the purpose of bringing some balance in the policymaking process of privacy legislation. In fact, privacy impact assessments (PIA) conducted by both public and private entities, often fail to bring the mentioned benefits into account. Therefore, considering that privacy risks, in general, have to be balanced against non-privacy rewards we need to try to list and categorize them in order to sketch a formula, that still lacks, to work out the balance (many mechanisms exist, on the other side, to assess privacy risks[36]).

4.1 Healthcare

Big data use in health care can provoke positive innovation from the first stages of prevention and identification of illnesses, to the treatment stage, allowing the analysts to study a myriad of interactions between a large number of different factors and drugs. The study of this enormous number of interactions would not be possible (and was not possible) in a small data world.

A government that launched initiatives to exploit the advantages of big data analytics within the medical field is the English one. David Cameron has recently announced[37] that every NHS patient would become a "research patient" whose medical records would be open up for medical research. The system chosen by the UK government

---

[35] E. Brynjolfsson et al., *Strength in Numbers: How Does Data-Driven Decision-Making Affect Firm Performance?*, A51 (2011), available at http://www.a51.nl/storage/pdf/SSRN_id1819486.pdf

[36] R. Clarke, *An Evaluation of Privacy Impact Assessment Guidance Documents*, Int'l data privacy law (2011). Available at http://idpl.oxfordjournals.org/content/early/2011/02/15/idpl.ipr002.full.pdf

[37] BBC, *Everyone 'to Be Research Patient', Says David Cameron, BBC UK Politics, Dec. 5, 2011*, http://www.bbc.com/news/uk-16026827

in this case to leave a choice to the public was an opt-out right. Cameron swore that "[this] does not threaten privacy […] it does mean using anonymous data to make new medical breakthroughs"; considering the "threat to the privacy" some legitimate doubts may arise, it is only matter of assessing if, again, it is a cost it is worth to be borne.

4.2 Smart Grids

The "smart grid" project illustrates how an advanced data analysis can improve the quality of living and decrease the costs. In a simple concept the smart grid is designed to permit users, service provider and other third parties to monitor the use of electricity.

Issues on the line, as outages of power, cyber attacks or natural disasters can be easily spotted and action can be taken. Moreover, the service provider could also use the new smart system in order to decide when and whether to switch from baseload to peak power plants.

On the privacy-side of the barricade, it is really interesting to notice how certain legal system; the Canadian in particular already faced a number of legal issues[38] and actual lawsuits related to the consequences of monitoring the electric usage of a private dwelling[39]. Furthermore, in the state of Ontario an entire report about the relation between the implementation of a smart grid system and the necessary PIA (Privacy Impact Assessment) has been drafted back in 2009[40].

4.3 Traffic Management and Mobile

Urban planners and governments around the world are looking at the analysis of personal location data in order to organize the movement of people in the cities, in this way is possible to cut the level of pollutants in the atmosphere and to fight congestion on the streets.

---

[38] A. Maykuth, *Utilities' Smart Meters Save Money, but Erode Privacy*, The Philadelphia Inquirer (2009).

see also A. Jamieson, *Smart Meters Could Be 'Spy in the Home',* The Telegraph, Oct. 11, 2009, available at http://www.telegraph.co.uk/finance/newsbysector/energy/6292809/Smart-meters-could-be-spy-in-the-home.html

[39] R. v. Gomboc, 2009 ABCA 276 at 17.

[40] *Smart Privacy for the Smart Grid: Embedding Privacy into the Design of Electricity Conservation* (Information and privacy commissioner of Ontario & future of privacy forum 2009).

Available at: https://www.ipc.on.ca/images/resources/pbd-smartpriv-smartgrid.pdf

At the same time individual drivers can benefit from real-time traffic information, accident reports and scheduled road works.

Moreover, for what concerns mobile technologies, at MIT and Harvard researchers are working on using mobile gathered information to predict food shortages, to quantify crime waves[41] and to assess how to intervene to improve the education system of the developing countries[42].

4.4 Retail and Payments

The big data trend[43] for organizations operating in the retail market is to make use of the analytics in order to link the online activity of the customers to the offline behavior in order to assess the effectiveness of their tailored ads according to in-store past purchases, and to retarget in-store customers with ads once online.

Another valuable use of big data is fraud detection in the online payment industry. Some companies have developed a form of predictive fraud score. In a few words the system tracks back the use of a card and gives it a score expressed in single number. In this way it is possible to achieve the measure of the likelihood of the fraudulence of a transaction

---

[41] J. Toole, *Quantifying Crime Waves*, Proceedings of Aaai Artificial Intelligence for Development (2010).

[42] M. Moussavi, *A Model for Quality of Schooling*, N. McGinn, Proceedings of Aaai Artificial Intelligence for Development. (2010).

[43] Opinion 2/2010 on Online Behavioral Advertising , at 5, WP 171 (Article 29 Working Party 2010).

5. Big Data and Open data: a synergy

Open data is the idea that certain categories of data should be freely available for anyone to use and republish with no legal impairments from copyright, patents or other legal constructs[44]. This free source of data can consist in the fuel to power the engines of thousand of startups, which apply the big data practices to extract value[45]. The major sources of open data project are Open Science Data and Open Government Data. Focusing on the latter, the trend of governments around the world launching countless open data initiatives is lead by a fundamental consideration[46].

Governments' position allows them to collect huge troves of data because, in contrast with data holders in the private sector, because they have the power to compel people to provide them with information while the commercial operators have to persuade them to provide data as consideration for an offered service. There are both legal and economic solid arguments in supporting the implementation of an open data policy[47]: some scholars argue that providing government information to the public in a machine-readable format may enhance government accountability, transparency and public participation. Other authors hold that the opening up of official information could support technological development and innovation, boosting economic growth enabling third parties to develop new applications for the datasets.

5.1 Public Sector Information Directive

The trend of opening up government data is continuously growing in the European Union too; the European Commission in its 2010 Digital Agenda particularly emphasized the idea of making available public sector information in order to incentivize markets for

---

[44] J. Brito et al., *Crowdsourcing Government Transparency*, Sci. & Tech. L. Rev. (2008).

[45] See www.flyontime.us as a successful example.

[46] I.a. *Executive Order 13642: Making Open and Machine Readable the New Default for Government Information* (US President Barack Obama), May 2013. Available at https://www.whitehouse.gov/the-press-office/2013/05/09/executive-order-making-open-and-machine-readable-new-default-government-

[47] J. Grey, *Towards a Genealogy of Open Data.*, General Conference of the European Consortium for Political Research. (2014). Available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2605828

See also note 41.

online content. The treasure chest of public sector data is a consistent one, with an estimated value of €27 billion.[48] The legal tool supposed to obtain such a treasure was the PSI (Public Sector Information) directive; the other factor which played a role in the genesis of the PSI directive were the concern of the EU commission of being under competitive with respect to the US.[49]

The directive was built on two main goals: on the one hand, enabling the availability of public sector information to third parties at low prices and competitive conditions, and on the other hand, ensuring a level playing field between public bodies that operate in the information market in competition with the private information industry.[50] The recognition of the added value for the whole society through the collection and reuse of public sector information operated by private actors under the PSI directive and its direct relation to big data model has been achieved. However, the main target group of the directive, was the information industry, it was "the creation of Community-wide information products and services based on public sector documents".

The latter goal seems still to be achieved due to the fact that it is not entirely clear and reflected in the practices of the single Member States.

5.2 Green and Blue Botton

In 2012 the Obama administration in the U.S. created the green button initiative: families and business would have given full access – in a computer-friendly and consumer-friendly format – to their energy usage information[51]. The administration put emphasis on the fact that the possibility of reusing these data would have strongly fostered new

---

[48] M. Dekkers, *Measuring European Public Sector Information Resources*, Mepsir (2006). Available at *www.ec.europa.eu/newsroom/document.cfm?doc_id=1198*

[49] K. Janssen, *Towards a European Framework for the Re-Use of Public Sector Information: A Long and Winding Road*, 11 International Journal of Law and Information Technology p.184-201 (Oxford University Press 2003).

[50] K. Janssen, *The Influence of the PSI Directive on Open Government Data: An Overview of Recent Developments*, 28 Government Information Quarterly 446 (Elsevier 2011).

Available at http://dx.doi.org/10.1016/j.giq.2011.01.004

[51] A. Chopra, *Green Button: Providing Consumers with Acces to Their Energy Data.*, Office of Science and Technology Policy Blog (2012). Available at

https://www.whitehouse.gov/blog/2012/01/18/green-button-providing-consumers-access-their-energy-data

technologies like management systems and smartphone applications able to interpret and use such information[52].

The Green Button project follows a path charted by another similar initiative taken in the field of health. This initiative was the Blue Button, launched in 2010 to give U.S. veterans the possibility of downloading their health records. Since than, the support of more than five hundreds private companies pursued the goal of increasing the patient access to their health data from health care providers, medical laboratories and retail pharmacy[53]. The return for the developers is huge amasses of useful information (immunizations, allergies, medications, family health history etc.) to aggregate[54].

5.3 Data.gov

An additional program wanted by the American government is the data.gov initiative (and its replicas all over in the world). The rationale behind it is that Governments have long been the biggest generators, collectors, and users of data (not necessarily PII), keeping records on every birth, marriage, and death, recording information and statistical models on all aspects of the economy, and maintaining statistics on licenses, laws, and the weather.

Until the last few years, data were difficult to locate and process even if publicly available[55]. The problem to deal with is that in many countries the freedom of having access to public data, about the social security system for example, would entail, at best, a

---

[52] A. Chopra, *Modeling a Green Energy Challenge after a Blue Button,* The White House Office of Science and Technology Policy (2011). Available at

https://www.whitehouse.gov/blog/2011/09/15/modeling-green-energy-challenge-after-blue-button

[53] Executive Office of the President, *Big Data: Seizing Opportunities, Preserving Values* (The White House reports 2014). Available at https://www.whitehouse.gov/sites/default/files/docs/big_data_privacy_report_may_1_2014.pdf

[54] A. Chopra et al., Blue Button' Provides Access to Downloadable Personal Health Data (White house Blog 2010). Available at https://www.whitehouse.gov/blog/2010/10/07/blue-button-provides-access-downloadable-personal-health-data.

[55] A. Conley et al., *Sustaining Privacy and Open Justice in the Transition to Online Court Records: A Multidisciplinary Inquiry*, 71 MD. L. REV. 772 (2012).

huge PDF document, difficult to explore and locked for editing. The stated purpose of the new website is to increase public access to high value, machine-readable datasets generated by the Executive Branch of the Federal Government.[56]

This huge opening up of governmental data should, in the intention of the promoters, unleash a wave of innovation and foster the creation of new economic value, as new businesses and individuals will start to use to raw data in order to improve their own existing services or to offer new solutions [57]

[56] The Open Society: Governments Are Letting in the Light, The Economist (2010). Available at http://www.economist.com/node/15557477

[57] *Id.*

6. The dark side of Big Data

The age of Big Data brings with it three main concerns: dangerous collateral effects of the need of data and usefulness of the results of their analysis. The following are the equivalent of the greenhouse gases consequent to the advent of the industrial age. We would never waive the innovations of the past 150 years and in the same way we will not renounce to the benefits of the big data era but we must be aware of the risks

The privacy is the first, most obvious candidate, to be targeted as at stake. Quality and quantity of information collected on a daily basis on our lives is growing in exponentially[58], the constant surveillance so dreamed by totalitarian regimes would be far easier today that we ought to hold fast to our freedoms.

The second risk posed by the shift of scale is the adverse effect of decisions and measures taken considering correlation instead of causation and so probability instead of certainty. The pre-criminal surveillance and imprisonment is just one out of many possible grotesque effects of it, applied for instance to genetic sequencing could be part of eugenic discriminatory politics based on propensity.

The third problem which could arise has been called "dictatorship of data"; in other words, the emphasized and fetishized use of the information within the decision making process could result in deep distortions and have, after all, a failing future.

6.1 The incremental adverse effect on Privacy

Not every application of big data analysis poses a threat to privacy; there are tons of uses that simply do not imply the collection of any information classified as personal. In all these cases a risk for privacy does not exist.

The actual threat to our personal information has the same characteristics of Big Data itself, this means that the dimension of the threat shifted its nature. In clearer words, the big data age is not only levering the degree of risk its applications can generate, in that case the new threat would be the same in nature but higher in intensity. Hence, in that case the solutions would arise from a narrower application of the prevailing laws and regulations on data and privacy protection; one should "only" redouble the quantity of a qualitatively identical effort.

---

[58] E. Quinn, *Smart Metering and Privacy: Existing Law and Competing Policies* (Colorado Public Utility Commission. 2009). Available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1462285

The problem is that the shift in quantity is resulting in a consequent shift in quality, the value of the information no longer resides in its first application, as we held, the reuse has become the core business for all the big data companies. Furthermore, the vast accumulation of personal data has an incremental adverse effect on privacy (also called "aggregation"[59]); in other words, the threat to individual privacy is more than proportional to the amount of data collected, moreover, once data are collected to an identified subject, they become very difficult to separate[60].

Likewise, the solutions designed by the policymakers have to follow a trend of qualitative change; three of the main techniques of protection (the "purpose specification" principle, the "use limitation" principle and the "notice and consent" model) fail to ensure an adequate level of protection for the following reasons.

The "purpose specification" and the " use limitation" principles find their origin in the first generation of data protection regulations with the clear purpose of lessen the risks of an extensive data collection and their unwanted consequence of surveillance and control[61].

During the era of the 80's and 90's privacy regulations these two principles played a key role in the genesis of the "notice and consent" model as the boundaries of the possible use of the data performed by the data controller[62].

The "notice and consent" model has become the cornerstone of privacy principles in the world. Today the individuals are told which information is being gathered and for what purpose; however, in a big data environment it is impossible to provide the user with

---

[59] O. Tene and S. Polonetsky, *Big Data for All: Privacy and User Control in the Age of Analytics*, Northwestern Journal of Technology and Intellectual Property, 11(239). (2013). See also D. Solove, *Access and Aggregation: Public Records, Privacy and the Constitution*, 86 Minn. L. Rev. 1137, 1160-64 (2002); and; D. Solove, A Taxonomy of Privacy, 154 University of Pennsylvania Law Review 477 (2006).

[60] Myspace LLC, F.T.C. File No. 102-3058, Agreement Containing Consent Order (May 8, 2012), (The charge was the "constructively share" personally identifiable information with third party advertisers by sharing with such advertisers a unique identifier assigned to the profile of each Myspace user (a "Friend ID"), which could then be used to access such user's profile information - a practice referred to in the industry as "cookie syncing).

[61] A. Mantelero, *The Future of Consumer Data Protection in the E.U. Re-Thinking the 'notice and Consent' Paradigm in the New Era of Predictive Analytics.*, Computer Law & Science Review, 652-652. (2014).

[62] Id.

a precise description for uses that are hidden in the folds of datasets[63]. Trying to imagine how to adapt the concept to the new contest it could be useful to think at the reuse clause in two opposite ways: the first with the higher protection possible for the individuals, the second with the higher incentives for innovation.

In the first case, with a strict application of the purpose specification principle, a company should be obliged to contact back every single user in order to obtain their consent for the new use; the other way round, asking the individuals to agree to any future possible use of their data would completely empty the meaning of "notice and consent". Applying the first model would entail, assuming the material feasibility, unbearable costs for the companies, impeding (or hugely lessening) big data-related benefits; applying the second would simply erase every shelter for individuals' privacy[64].

Other methods of protection could fail either if considered as a valuable on their own; for example, if everyone's information has become part of a dataset, even opt-out may leave a trace.

Even anonymization does not effectively work in many cases: two researchers of the University of Texas re-associated de-identified Netflix movie recommendations with identified individuals by crossing a de-identified database with publicly available resources accessible online[65]. The two scholars explained the phenomenon holding: "Once any piece of data has been linked to a person's real identity, any association between this data and a virtual identity breaks anonymity of the latter".

A dark picture of a "database of ruin" has been also drawn in relation of the consequences of the incremental effect as depicted in the lines above[66].

Moreover, considering the current landscape of the big data actors and their considerable power in the market, a number of scholars argued that it is the end of the user's self-determination as we know: nowadays individuals are either unable to understand data

---

[63] Article 29 Data ProtectionWorking Party, 'Opinion 06/2013 on open data and public sector information ('PSI') reuse' (2013) 19e20 and Article 29 Data Protection Working Party, 'Opinion 03/ 2013 on purpose limitation' (2013).

[64] See note 5.

[65] A. Narayanan, Robust De-Anonymization of Large Sparse Datasets, V. Shmatikov, Iee Symp. On Security & Privacy 111 (2008).

[66] P. Ohm, Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization, 57 Ucla L. Rev. 1701, 1748 (2010).

processing and its purposes or they are not in a position that gives them the possibility to decide freely ("take it or leave it" agreements)[67].

The search and need for legislative solutions to contrast the weakness of the consolidated methods o protection is ongoing with the draft of the new General Data Protection Regulation (GDPR) in Europe[68] and with the rethinking of the legal framework in the U.S.[69], meanwhile the Canadian common law is rapidly evolving to face the challenges posed by the consequences of misuses of personal information[70].

6.2 Predictive analysis – probability and punishment

Big data may facilitate enormously predictive analysis with stark implications for all those categories of subjects susceptible to be prone to disease, crime or other socially stigmatizing behaviors or traits. The seeds of the fictional pre-crime policy depicted in Minority Report have already been sowed; in more than half of all the U.S. parole boards use predictions founded on data analysis as a valuable factor to decide if someone has to remain in prison or can be released.

Although predictive analysis might be useful in certain fields as law enforcement, national security, credit screening it could also enhance illegal activities such as "redlining"[71], nevertheless it raises concerns, for instance in the law enforcement arena, about surveying or even incarcerating suspects based on elements that are more similar to thoughts than deeds[72]. This kind of enforcement, clearly unconstitutional under European

---

[67] The Boston Consulting Group, The value of our digital identity (2012), available at http://www.libertyglobal.com/PDF/public-policy/The-Value-of-Our-Digital-Identity.pdf.

D. Solove, *Privacy Self-Management and the Consent Dilemma*, 26 Harvard Law Review 1880 (2013).

[68] Commission proposes a comprehensive reform of data protection rules to increase users' control of their data and to cut costs for businesses. (2012) *European Commission Press Release.*

[69] See note 12.

[70] Jones v. Tsige, 2012 ONCA 32, 2012-01-18 see also R. Barrass and L. Wasser, Seclusion Intrusion: A Common Law Tort for Invasion of Privacy, McMillan LLP (2012).

[71] The practice (in the U.S.) of denying services, either directly or through selectively raising prices, to residents of certain areas based on the racial or ethnic makeups of those areas.

[72] R. Van Brakel and P. De Hert,) Policing, Surveillance and Law in a Pre-Crime Society: Understanding the Consequences of Technology Based Strategies, 20 J. Police Stud. 163 (2011).

and North American constitutions, could become "popular" under totalitarian regimes in the world, or in the U.S. too in case of and extensive use of emergency legislation[73].

Moreover, every typology of predictive analysis blind to outliers that applies the biblical "whbwwb" (what has been is what will be) rule will become guilty of being self-fulfilling prophecies that accentuate the social stratification[74].

In all the above-mentioned hypothesizes the adverse effects of predictive analysis systems is strictly related to a misuse of PII, what if it would be possible to draw sensible predictions from non sensible data through their analysis? The case is a famous one[75]: the main character is the U.S. retailing giant, Target Inc, its statisticians created a "pregnancy prediction score" based on the historical buying records of women who had signed up for baby registries. The employed statisticians were able to sift out a set of products that, when grouped, gave Target the chance to accurately predict a customer's pregnancy and due's date with a minimum margin of error. The other character of the case was the father of a teenage girl who strongly complained that his daughter received advertisements for baby products. A few days later, he called the store to apologize admitting that, "There's been some activities in my house I haven't been completely aware of. She's due in August."

In this case Target had never collected personal data, however, the prediction constituted a personally sensitive information and Target also reused this predictive PII for marketing purposes.

6.3 Is Competition at stake?

The concerns about the possibility of Big Data to represent a durable entry barrier for online services, with the consequent entrenchment of large online firms, arose from recent calls for antitrust intervention; in particular when big data consist in personal information. According to the data protection supervisor, for instance, in the case of the acquisition of Dataclick performed by Google in 2007[76], the approach of the EU Commission was too economic in assessing the antitrust relevance of the operation, not considering how the acquisition could have influenced the users whose data would be further processed by merging the two companies' datasets, conceivably to provide services,

---

[73] M. Rotenberg, Foreword: Privacy and Secrecy after September 11, 86 Minn. Law Rev. 1115 (2002).

[74] J. Stanley, Eight Problems with "Big Data, Aclu Blog (2012).

[75] C. Duhigg, How Companies Learn Your Secrets, N.Y. Times Magazine (2012) available at http://www.nytimes.com/2012/02/19/magazine/shopping-habits.html?_r=0

[76] L. Story and M. Helft, Google Buys DoubleClick for $3.1 Billion, N.Y. Times Magazine (2007) http://www.nytimes.com/2007/04/14/technology/14DoubleClick.html

perhaps bundled or even tied to the simple search service, that were not envisaged when the data were originally submitted[77]. Authors in the U.S expressed some doubts about a deregulated landscape for big firms operating in the data market (whose existence has been challenged too)[78].

Some scholars[79] believe that Big Data is the next big trend in antitrust law while others[80] simply hold that big data is neither a "product" in the antitrust lexicon nor a typology of input by business need to obtain from outside its contest in order to compete. The author's argument goes on justifying his thesis (on the nature of data as an antitrust issue) holding that the existence of a data market and of a consequent "market power" is highly implausible given the ubiquitous and non-rivalrous nature of information.

The arguments listed by the supporters of the opposite thesis to deny the concept of data neutrality from a competition angle are several:

- many online companies have adopted business models that rely on personal data as a key input;
- the same number of companies undertake data-driven strategies in order to obtain and sustain competitive advantages;
- the battle over personal data has spread to strategic acquisitions as reported by the OECD (Organisation for Economic Co-operation and Development);
- if data-driven businesses bear significant costs to obtain, store, and analyze data probably they could have strong incentives to impede their competitors' access to these datasets;

---

[77] Eur. Data Prot. Supervisor, Privacy and Competitiveness in the Age of Big Data: The Interplay between Data Protection, Competition Law and Consumer Protection in the Digital Economy., Preliminary opinion (2014). Available at https://secure.edps.europa.eu/EDPSWEB/webdav/shared/Documents/Consultation/Opinions/2014/14-03-26_competitition_law_big_data_EN.pdf

[78] N. Newman, Search, Antitrust and the Economics of the Control of User Data, 31 Yale J. On Reg. 401 (2014).

[79] A. Grunes and M. Stucke, *No Mistake about It: The Important Role of Antitrust in the Era of Big Data.*, American Bar Association ABA: Antitrust Source. (2015).

[80] D. Tucker, *Big Mistakes Regarding Big Data*, H. Wellford, American Bar Association ABA: Antitrust Source. (2015).

- companies, whose business model strongly depends on securing a competitive advantage through big data, could also undertake anticompetitive data-driven strategies.

The subject matter presents elements of complexity that cannot be fully examined here. However, it is important to consider the presence of antitrust and competition law concerns dealing with privacy and data protection regulation that may have an impact.

6.4 In the realm of data barons

A famous adagio goes: "if you are not paying for it, you are not the customer, you are the product[81]"; the argument is that an unbalanced situation of redistribution of the benefits coming from big data exploitation exists.

In first place the previous statement is grounded on the fact that online interactions are more similar to barter-like transactions with individuals giving away personal information for free services[82].

Moreover, these transactions appear to take place in an inefficient market, characterized by steep forms of information asymmetry, Big Data is boosting this effect and there is more; a vast use of data driven price determination based on lifestyle habits and personal preferences can easily result in chewing, a bit after the other, the entire value surplus available for the customers. How? Setting the price of goods and services as close as possible to the individual's reservation price.

Data collectors have been surely enriched by the data deluge and have enriched the individuals with the availability of a huge number of services that would not be thinkable outside a Big Data society.

On the other hand, legislators should verify a balanced redistribution of the wealth, avoiding too tight provisions that would hamper and suffocate the incentives for economic actors and too loose provisions susceptible of being scarcely effective.

---

[81] J. Zittrain, *Meme Patrol: When Something Online is Free, You 'Re Not the Customer, You 'Re the Product*, The Future of The Internet (2012). Available at http://blogs.law.harvard.edu/futureoftheinternet/2012/03/21/meme-patrol-when-something-online-is-free-youre-not-the-customer-youre-the-product/

[82] C. Anderson, Free: The Future of a Radical Price (Hyperion 2009).

II. Individual privacy and data protection in the era of big data

1. The evolution of the United States privacy legal framework

1.1 The Origins

The roots of the modern conception of privacy (but the plural should be more appropriate[83]) come from the seminal and renown article[84] of Samuel Warren and Louis Brandeis "The right to Privacy", in which they argue that:

"[r]ecent inventions and business methods call attention to the next step which must be taken for the protection of the person and for securing to the individual… the right 'to be let alone'… numerous mechanical devices threaten to make good the prediction that 'what is whispered in the closet shall be proclaimed from the house-tops."

This work and its clairvoyance stand as one of the founding pillars of the evolution[85] of the privacy US common law, undertaken in the XX century, building up the principles for the protection of citizens' right to privacy from both the government and other individuals[86].

It is in the Constitution that this "new" right finds its justification, in particular the Fourth Amendment ensures that "the right of the people to be secure in their persons, houses, papers, and effects, against unreasonable searches and seizures, shall not be violated, and no warrants shall issue, but upon probable cause, supported by oath or affirmation, and particularly describing the place to be searched, and the persons or things to be seized" in order to limit the possibility of a government's intrusion upon individuals' private life.

---

[83] See note 60, (A taxonomy of Privacy).

[84] S.D. Warren and L.D. Brandeis, *The Right to Privacy*, 4 Harvard Law Review 193 (JSTOR 1890).

[85] According to legal scholar Roscoe Pound, the article did "nothing less than add a chapter to our law", and in 1966 legal textbook author, Harry Kalven, hailed it as the "most influential law review article of all". As recently as 2001, in the Supreme Court case of Kyllo v. United States, 533 U.S. 27 (2001), the article was cited by a majority of justices, both those concurring and those dissenting.

[86] W. L. Prosser, *Privacy*, 48 California Law Review 383 (California Law Review 1960).

Time and technology have shaped, over the course of the last century, what case law considers as a "search" for the purposes of the fourth amendment. The Supreme Court gave its contribution since the 1928 famous case *Olmstead v United States,* in which the majority of the justices held that placing wiretaps on a phone line located outside of a person's house did not violate the Fourth Amendment, even though the government obtained the content from discussions inside the home. However, the decision went down in history not for the holding[87] but for the dissenting opinion drafted by Justice Brandeis who enshrined the right to privacy, conferred by the framers, defining the right to be let alone – "the most comprehensive of rights and the right most favored by civilized men".[88]

The Court's opinion in *Olmstead* remained the law of the land until year 1967, in which it was overturned by the court's decision in *Katz v. United States.*

In Katz, the Court held that the FBI's placement of a device for recording on the outer surface of a public telephone booth without a warrant, even though the device did not physically entered the booth, his person, or his property qualified as a search that violated the "reasonable expectation of privacy" of the person using the booth.[89] The juridical norm in *Katz* clearly stated that an individual's subjective expectations of privacy are protected when society regards them as reasonable.[90]

The reaction of civil courts was definitely slower; they did not recognize privacy as a valid cause of action to sue. One had to wait until 1934 when Restatement of Torts recognized an "unreasonable and serious" invasion of privacy as a basis to bring legal action[91]. The courts began considering privacy as a valid cause of action, however from the work of the courts sprang not one but a complex of four different potential torts:

1. intrusion upon a person's seclusion or solitude, or into his private affairs[92];
2. public disclosure of embarrassing private facts about an individual;

---

[87] Olmstead v. United States, 277 U.S. 438 (1928).

[88] Id. at 478.

[89] Katz v. United States, 389 U.S. 347, 361 (1967) (Harlan, J., concurring).

[90] W. Lafave, *Search and Seizure: A Treatise On The Fourth Amendment*, §§ 1.1–1.2 West Publishing, 5th ed (2011). ("[L]ower courts attempting to interpret and apply Katz quickly came to rely upon the Harlan elaboration, as ultimately did a majority of the Supreme Court").

[91] Restatement (First) Torts § 867 (1934).

[92] See note 94; see also Restatement (Second) Torts § 652A (1977) (Prosser's privacy torts incorporated into the Restatement).

3. publicity placing one in a false light in the public eye;
4. appropriation of one's likeness for the advantage of another

Some scholars argued that the protection provided by the complex of four was insufficient to deal with privacy issue arising from intensive and massive collection, use and disclosure of personal information by businesses in the modern marketplace[93]. The same idea of inadequacy was expressed in relation to the rise of automated processing methods of data crunching.

1.2 The rise of the Fair Information Practice Principles

In 1973 the U.S. Department of Health, Education, and Welfare issued a report entitled "Records, Computers, and the Rights of Citizens"[94]. The content of the report was an analysis of the harmful consequences of automated data processing on individuals' right to privacy. The rules and safeguards elaborated in that occasion became known as Fair Information Practice Principles (FIPPs) and represent the classic framework of the modern protection regimes.

The FIPPs are , in their essence, a set of principles to provide basic protections for handling personal data. The main principles are:

1. *notice/awareness* (consumers should be given notice of an entity's information practices before any personal information is collected from them);
2. *choice/consent* (the idea is to give consumers the possibility to control how their data is used[95]);
3. *access/participation* (access does not only imply the possibility for the user to view the data collected in an inexpensive and timely way, but also to verify and contest its accuracy);
4. *integrity/security* (the organization that collects information has an obligation to ensure that the data is reliable and kept secure);

---

[93] i.a. K. Taipale, *Data Mining and Domestic Security: Connecting the Dots to Make Sense of Data*, V The Columbia Science and Technology Review (2003).

[94] Pub. L. 93-579 (codified at 5 U.S.C. § 552a).

[95] for a critique of the principle e.g. in the field of healthcare see:
H. Tavani, The Consent Process in Medical Research Involving DNA Databanks: Some Ethical Implications and Challenges, M. Bottis, ACM SIGCAS Computers and Society, 40(2), 11-21 (2010).

5. *enforcement/redress* (enforcement measures have to be enacted in order to guarantee that companies duly follow the principles; these could be self regulation by the information collector, private remedies that give causes of action for individuals, government enforcement),

These principles constitute the basis for the draft of the 1974 Privacy Act, which provides a regulatory environment for maintenance, collection, use and dissemination of personal information performed by the federal government in systems of records[96].

In 2012 the White House outlined and released its version of the FIPPs wih the attempt of widening the scope and enhancing the focus on data's user and categorization and individual control over the information[97].

Yet even these broadened principles imply the feature of depending not only on the knowledge of which information is considered personal data but also on the necessity of providing notice, choice, and control to users ex ante any privacy harm. Also in this modern version of the old trusted principles Privacy law is primarily concerned with causality, whereas Big Data, as we noticed before, is generally a tool of correlation[98].

1.3 The current landscape of privacy protection in the United States

The philosophical background of the patchwork quilt of the US privacy protection system – made up of common law, federal legislation, the US constitution, state law and certain state constitutions[99] – can be summarized in its essence by the triptych of values:

---

[96] 88 Stat. 1896. Pub. Law 93-579 and

Organization for Economic Cooperation and Development, *Thirty Years After The OECD Privacy Guidelines p.17* (2011), available at http://www.oecd.org/sti/ieconomy/49710223.pdf

[97]Information is categorized according to its information type. An information type is a specific category of information (e.g., privacy, medical, proprietary, financial, investigative, contractor sensitive, security management) defined by anorganization or, in some instances, by a specific law, Executive Order, directive, policy, or regulation; see also The White House, Consumer data privacy in a networked world: a frame work for protecting privacy and promoting innovation in the global digital economy (2012), available at https://www.whitehouse.gov/sites/default/files/privacy-final.pdf.

[98] I.S. Rubinstein, *Big Data: The End of Privacy or a New Beginning?*, 3 International Data Privacy Law 74 (Oxford University Press (OUP) 2013).

[99] A. Levin and M.J. Nicholson, *Privacy Law in the United States, the EU and Canada: The Allure of the Middle Ground*, University of Ottawa law & Technology Journal 357(2005). Available at http://uoltj.ca/articles/vol2.2/2005.2.2.uoltj.Levin.357-395.pdf

"life, liberty and the pursuit of happiness"[100]; it expresses the desire of the American people to be let alone by its government. The aim is to employ a sectorial approach that focuses on regulating specific risks of privacy harm in particular contests, as health care and credit. This approach places fewer wider rules on the use, collection, re-use and dissemination of data in order to boost innovation coming from private entities, the drawback in doing this is the risk of leaving unregulated some potential typologies of exploitation of data in regions that fall between two regulated sectors.

It is interesting to consider, incidentally (the point will be further discussed), a parallelism with the Canadian constitutional values and consequent enactment through legislative measures.

The Canadian belief in "peace, order and good government"[101] brought their legal conception of privacy toward shores more similar to the European Union ones, conceiving privacy as fundamental human right**,** generally involving top-down regulation. Moreover, the Canadian triptych led them to a middle ground between US and EU: sharing with the former the fears of a Big Brother government and with the latter big concerns about the possible abuse of personal information performed by the private sector.

One of the issues of the disjointed legal framework for privacy of the US is that common law and constitutional protection overlap for informational privacy. The outcome of this unpredictability is the consequent difficulty to articulate any kind of overall legal theory with respect to privacy[102].

If the protection provided by the constitution and its enactment against government intrusion seems so far achieved, the same cannot be hold with respect to disclosure of personal matters[103].

Statutes have filled in the insufficiencies of common and constitutional law, but a number of narrow, specific state and federal statutes addressing the protection of PII tackle specific issues rather than covering the category of privacy as a whole.

In some cases, state legislators have been more conscious of the need of an up-to-date legal framework to adapt the legal system to the evolution of technology; however, their effect is limited due to jurisdictional limitations and states' weak enforcement

---

[100] United States, Declaration of Independence (1776).

[101] Constitution Act, 1867 (U.K.) 30 & 31 Vict. C.3, s. 91, reprinted in R.S.C. 1985, App. II, No. 5.

[102] D. Solove, *Conceptualizing Privacy*, 90 California Law Review 10 (2002).

[103]W. DeVries, *Protecting Privacy in the Digital Age*, 18 Berkeley Technology Law Journal 283 p.285 (2003).

see also Whalen v. Roe, 429 U.S.589, in which the Court refused to find that the government's recording of personal drug prescription information violated the constitutional right to privacy because the information was adequately protected.

capabilities[104]. On the other hand, the reactive, adaptive process adopted by the courts has adverse consequences on the way digital privacy problems are addressed rationally and effectively.

1.4 Consumer Privacy Bill of Rights

In 2012 the U.S. government issued its "Privacy Blueprint" in order to address two main issues affecting the current framework: a clear statement of basic privacy principles that apply to the commercial world and a sustained commitment of all stakeholders to address consumer data privacy issues as they arise from advances in technologies and evolution of the business models[105].

This report is made up of of four main elements in order to specify and give execution to the main goals of the administration: first, a Consumer Privacy Bill of Rights based on the Fair Information Practice Principles, second, a call for government-convened multi-stakeholder processes to apply those principles, third, a plan for an effective enforcement of privacy-related rights and fourth, a commitment of adherence to international privacy regimes that allow the cross-border data flow.

Most of the rights of the Bill find their logical origin in the FIPPs with some variation wanted to adapt the old principles to the new challenges. The listed rights are:

i. *Individual control*, consumers must be granted the right to exercise control over the collection and use that organizations make of their data,

ii. *Transparency,* the information concerning privacy and security practices must be provided in a clear and easily understandable way,

iii. *Respect For Context,* is the right to expect that the collection, use and dissemination/disclosure performed by organizations would be performed consistently with the context in which the consumers provide the data,

iv. *Security,* consumers have a right to secure and responsible handling of their personal data,

---

[104] See note 111.

[105] White House Report, *Consumer Data Privacy in a Networked World: A Framework for Protecting Privacy and Promoting Innovation in the Global Digital Economy* (2012).

p.I, available at https://www.whitehouse.gov/sites/default/files/privacy-final.pdf

v. *Access and Accuracy,* consumers have a right to access their data at any time and in a usable format, they have also the right to correct the information believed inaccurate,

vi. *Focused Collection,* consumers have a right ask the data collectors to limit the scope of the collection according to the service provided,

vii. Accountability, consumers have a right to have personal data handled by companies with appropriate measures in place to assure their adherence to the principles listed within the Consumer Privacy Bill of Rights.

Compared to previous privacy frameworks, the Consumer Privacy Bill of Rights is more focused on the user with a more friendly and easier to understand lexicon[106].

Moreover, the rationale on which the CPBR is grounded is the idea of not requiring the organizations to comply with a strict set of rules and requirements; instead, the companies should have more freedom in determining how to implement the principles. Finally, the model conceived by the U.S. government, built up on the combination of broad baselines principles and specific codes of conduct can better protect consumers while supporting innovation[107].

1.5 The US Legal Framework in Action – The *actual* Consumer protection against re-use in the US Case Law

It is well established and generally recognized that end-users never read the terms of service agreements, as a result, the understanding of their ramification and a reasonable expectations of the dynamic collection, use and re-use are purely fictional[108].

Nevertheless, courts of law in the U.S. often uphold terms of service agreements in the digital world, with the consequence of keeping the current regime alive and well[109]. The consumers often invoke, albeit unsuccessfully most of the times, federal statutes, such as

---

[106] For example, it describes a right to "access and accuracy," instead of the previous formulations referencing "data quality and integrity." Similarly, it assures consumers that companies will respect the "context" in which data is collected and used, replacing the term "purpose specification".

[107] See note 12 p. 20.

[108] A. McDonald and L. Cranor, *The Cost of Reading Privacy Policies*, 4 I/S: J.L. & Pol'y 540 (2008).

[109] J. Moringiello and W. Reynolds, *From Lord Coke to Internet Privacy: The Past, Present, and Future of the Law of Electronic Contracting*, 72 Md. L. Rev. 452 (2013).

the Electronic Communication Privacy Act (ECPA)[110], Computer Fraud and Abuse Act[111], Video Protection Privacy Act[112], and state laws[113], in order to sue online-service providers for alleged privacy violations.

Moreover, the federal Wiretap Act[114], as amended by the ECPA, prohibits the unlawful interception of wire, oral, or electronic communications; consequently consumers have attempted to use this law and the related Stored Communications Act[115] to sue companies engaged in the online collection, use and sharing of data[116].

Now we will attempt to give a partial but coherent overview of this case law, trying to summarize the main categories of claims through famous and oft-cited cases. A seminal and famous case that involved the online ad data broker DoubleClicked posed the dividing line between cookies and other tracking activities illegal under the ECPA[117].

Beyond web browser tracking activities, plaintiffs have also tried to use the ECPA to restrict the commercial use of their data performed by ISPs. The famous case in this direction is *Kirch v. Embarq,* in which the 10th Circuit of the U.S. Court of Appeals held that an ISP is not infringing ECPA when it permits a third party to access the network traffic to conduct market research about the online behaviors of ISP's customers[118].

And there is more: the issues about the enforceability of the TOS is far to be solved, several recent cases out of Silicon Valley have showcased how difficult could be for users to face a victorious challenge of the terms of service and accompanying privacy policies[119].

---

[110] Electronic Communications Privacy Act of 1986, Pub. L. No. 99-508, 100 Stat. 1848 (1986) (codified as amended at 18 U.S.C. §§2510-32, 2701-12, 3121-27 (2013)).

[111] Computer Fraud and Abuse Act of 1986, Pub. L. No. 99-474, 100 Stat. 1213 (1986) (codified at 18 U.S.C. § 1030 (2013)).

[112] Video Privacy Protection Act of 1988, Pub. L. No. 100-618, 102 Stat. 3195 (1988) (codified at 18 U.S.C. § 2710 (2002)).

[113] E.g. the California Consumer Legal Remedies Act, Cal. Civ. Code §§1750- 56 (West 2013), and the California Comprehensive Computer Data Access and Fraud Act, Cal. Penal Code § 502 (West 2013).

[114] 18 U.S.C. §§2510-2522 (2013).

[115] Id. §§2702-2711.

[116] For an overview of several cases involving federal legislation and online privacy claims see J. Frieden et al., *Putting the Genie Back in the Bottle: Leveraging Private Enforcement to Improve Internet Privacy*, 37 Wm. Mitchell L. Rev. 1671 p.1706-1712 (2011).

[117] In re DoubleClick Inc. Privacy Litigation, 154 F. Supp. 2d 497 (S.D.N.Y. 2001).

[118] Kirch v. Embarq Mgmt. Co., 702 F.3d 1245, 1248-49 (10th Cir. 2012), cert. denied, 133 S. Ct. 2743 (2013).

[119] i.a. Swift v. Zynga Game Network, Inc., 805 F. Supp. 2d 904, 910-12 (N.D. Cal. 2011).

Another impairment for plaintiffs is that, according to the Northern District of California, proving injury in fact is a hurdle for plaintiffs who claim nothing more than the mere monetization of their information as their alleged harm[120]. This case made also clear that the ECPA might also be problematic for plaintiffs because the *ordinary course of business* exception to the ECPA's restrictions has been interpreted to include those furthering legitimate business purposes[121].

Despite significant court hurdles for plaintiffs who attempt to challenge terms after consenting to them, the FTC offers another approach in shaping privacy policies through its enforcement role against companies that engage in false and misleading practices[122]. Its approach, based on encouraging compliance to the industry standards on one hand and its higher specific weight in filing a complaint for an alleged breach of consumers' privacy, led big data subjects to settle the dispute several times.

Given this scenario, of companies able to lawfully make a robust use and share consumers' data under their privacy policy; it may be advantageous for companies to be somewhat vague in their terms of service and data use policies to shield full disclosure or not divulge trade secrets while still protecting them from potential liability.

---

[120] In re Google, Inc. Privacy Policy Litig., C-12-01382-PSG, 2013 WL 6248499 (N.D. Cal. Dec. 3, 2013).

[121] A. Bagley and J. Brown, *Limited Consumer Privacy Protections against the Layers of Big Data*, 31 Santa Clara Computer & High Tech. L.J. 483 (2014).

[122] J. Cox and K. Cline, *Parsing the Demographic: The Challenge of Balancing Online Behavioral Advertising and Consumer Privacy Considerations*, 15 J. Internet L. 1, 3(2012).

1.6 Interoperability and future goals of the American privacy legal system

In 2014 the FTC announced the beginning of a tighter cooperation with the representatives of the agencies of European Union and Asia-Pacific Economic Cooperation with joint EU and APEC endorsement of a brand new document drafted to map the requirements of the European and APEC privacy frameworks[123]. The project is supposed to help companies seeking to do business in both European and APEC countries, recognizing overlaps and gaps between the two legislations; in other words it should work as a tool to smooth the differences between the different regimes in order to facilitate commercial interoperability.

This kind of efforts clarifies obligations for companies and helps build interoperability between global privacy frameworks.

Trying to trace a common line of the story and goals of the American privacy legal system, we may argue that the U.S. government believes that the most common privacy risks still involve "small data"; considering that these risks usually do not involve especially large volumes, rapid velocities, or great varieties of information, nor do they implicate the kind of sophisticated analytics associated with big data, the policy that is being undertaken by the White House is to verify the resilience of the up-to-dated legal tools provided by the enactment of the privacy blueprint. In case of issues arising from the development of hazardous big data practices a commitment has been made of adapting the current legislation[124].

[123] W.P. 29, Press Release: "Promoting Cooperation on Data Transfer Systems Between Europe and the Asia-Pacific," March 26, 2013, available at http://ec.europa.eu/justice/data-protection/article-29/press-material/press-release/art29_press_material/20130326_pr_apec_en.pdf

[124] See note 12.

2. The current Canadian privacy legal framework

Compared to the US, the Canadian legal system is more recent and the first appearance of the concept of privacy was firstly made law by the British Columbia policymakers, not by the Ottawa's central government[125].

British Columbia's *Privacy Act* created a statutory tort for invasion of privacy in 1968[126]; later in 1982 the Canadian Charter of Rights and Freedoms enshrined the concept of privacy in the Canadian constitution[127]. The following year the *Privacy Act* became effective, regulating the ways the federal government used personal information[128]; then the Canadian citizens had to wait until 2000 to have the collection of personal information performed by private organizations regulated by the law with the *Personal Information Protection and Electronic Documents Act* (PIPEDA)[129]. In terms of the common law it was until 2012 that the Court of Appeal for Ontario recognized a tort for privacy invasion[130].

Thus, three main typologies of privacy claims can be made in Canada today, constitutional, regulatory and tort. We will attempt to provide an overview of the Canadian framework toward privacy analyzing the development of the three categories.

2.1 The constitutional claim

The Core of the constitutional claim is Section 8 of the *Charter of Rights and Freedoms,* providing that "[e]veryone has the right to be secure against unreasonable search or seizure"[131], its analysis and interpretation has to be conducted tracing its evolution through the trail of Supreme Court judgments.

The foundational case under this section is *Hunter v Southam*[132]; justice Dickson drew his opinion strongly relying on the American case *Katz* (*supra*), holding that privacy protected people not places, so Canadian scholars argued that the court did not really

---

[125] R.L.D.Hughes, *Two Concepts of Privacy*, 31 Computer Law & Security Law Review 4 *pp.527-537* (2015).

[126] S.B.C. 1968, c. 39.

[127] Canadian Charter of Rights and Freedoms, Part I of the Constitution Act, 1982, being Schedule B to the Canada Act 1982 (UK), 1982, c 11 (Charter of Rights and Freedoms).

[128] Privacy Act, R.S.C. 1985 c P-21.

[129] Personal Information Protection and Electronic Documents Act, SC 2000, c.5 (PIPEDA).

[130] *Jones v. Tsige* [2012] O.N.C.A. 32 [Jones].

[131] See note 123 section 8.

[132] Hunter v Southam [1984] 2 S.C.R. 145 at 155 [Hunter].

added any useful insight, limiting itself to the *Brandesian* conception of Privacy as a liberal right to be left alone by the government[133].

The Canadian People had to wait four years more in order to assist to an evolution of the meaning of privacy under section 8, it happened in *R v. Dyment*[134].

Warren and Brandeis article was not taken into consideration and the court strongly relied on a governmental report[135] as well as on the work of Alan Westin[136]. Summing up the court accepted the tridimensional theory of privacy[137] and the conception of privacy for the justices, while still underpinned by a strong notion of dignity, it encompasses more than the simple right to be left alone.

In *R. v Plant*[138] the notion was dried to a narrower 'biographical core of personal information' and since then became entrenched in Supreme Court's jurisprudence[139].

Later on, the Court tried to broaden the scope of the biological core of personal information[140] only to return on its footsteps few years later facing internet-related privacy breaches[141].

What is noticeable from the evolution over the years is that the Court's jurisprudence has forged a narrower and more instrumental understanding of privacy under s.8. One of the blames that can be addressed to the Supreme Court is to have never defined how and why privacy "goes to the essence of a democratic state". This statement

---

[133] See note 121.

[134] R. v. Dyment [1988] 2 S.C.R. 417 [Dyment].

The facts in that case were that Mr. Dyment was involved in a car accident and taken to hospital. While there, a doctor took a vial of Dyment's blood from an open wound. Dyment later disclosed to the doctor that he had a drink and the doctor turned the vial over to the police. The police analyzed the blood and charged Dyment with impaired driving. The question was whether the police needed a warrant to obtain the blood; was this a 'seizure' for the purposes of s. 8?

[135] Privacy and Computers, the Report of the Task Force Established by the Department of Communications/Department of Justice (1972).

[136] A. Westin, Privacy and Freedom (Scribner 1967).

[137] Territorial, bodily and informational.

[138] R v. Plant [1993] 3 S.C.R. 281 [Plant].

That case involved the police carrying out a 'perimeter search' of a house, as well as a check of the records of electricity consumption from that house, the combined results of which were then used to obtain a warrant that led to the discovery of a marijuana grow operation.

[139] See i.a. R v. Gomboc [2010] 3 S.C.R. 211 (supra), and R. v. Tessling [2004] 3 S.C.R. 432 [Tessling].

[140] *R. v. A. M.* [2008] 1 S.C.R. 569 [A.M.] and *R. v. Kang-Brown* [2008] 1 S.C.R. 456.

[141] *R. v. Cole* [2012] 3 S.C.R. 34 [Cole] and R. v. Vu [2013] 3 S.C.R. 657.

tends to suggest that, regardless to the fairly firm foundations in *Hunter* and *Dyment,* nowadays in Canada, when it comes to balancing between the individual interests in maintaining the control over his or her personal information and society's interest in effective policing, typically the latter will trump.

Eventually, the right to privacy under sec. 8 appears to be fairly weak considering that: the biological core is inherently an elastic concept not solid enough to shield the individual toward the claim of the public enforcement, also when admitting a breach of the individual's right to privacy the Supreme Court has been often prone to admit the use of the information as evidence under s.24 (2) nullifying the effectiveness of the "biological core" concept.

It has been skeptically noticed that if the Court truly believed that privacy was part of a person's inviolable personal essence or 'the right most valued by civilized men'[142] it would have judged the use of the information as evidence simply inconsistent with the Constitution[143].

2.2 The regulatory claim

The second way privacy is addressed in the Canadian legal system is in statutes whose aim is to regulate how both public and private organizations collect and make use of personal information.

As we have seen for the constitutional portrait of privacy, no definition can be found in statutes (state and federal) of what privacy is[144]. Their focus is mainly on what personal information use and its collection mean; they are also shaped on the OECD's eight privacy principles[145].

Furthermore, until a recent case[146], there had been little interplay between the broad wording of the statutes that sought to regulate the collection and use of recorded information about an identifiable individual and, on the other hand, the already analyzed growing recognition in the s.8 jurisprudence that what really mattered was the control over

---

[142] See *Olmstead* infra note 94.

[143] D. Stuart, *The Unfortunate Dilution of Section 8 Protection: Some Teeth Remain*, 25 Queens L. J. 65 (1999).

[144] in addition to the statutes cited above in the text, *see Freedom of Information and Protection of Privacy Act*, R.S.B.C. 1996 c. 165 [FIPPA]

[145] *OECD's privacy framework, OECD* (2013) available at http://www.oecd.org/sti/ieconomy/oecd_privacy_framework.pdf

[146] Alberta (Information and Privacy Commissioner) v. United Food and Commercial Workers, Local 401 [2013] 3 S.C.R. 733 [Local 401].

the information constituting the *biological core* of the individual. With one fell swoop the Court changed approach in *Local 401*.

The case, a classic hypothesis of freedom of expression vs right to privacy, however, differently from the jurisprudence of the ECtHR, in which the two rights are balanced one against the other according to the ECHR list of rights, the Canadian Supreme Court recognized that s.8 could not be support a claim of a privacy breach performed by a non public organization. And there is more, the defendant could challenge the constitutionality of the state provision[147] against s.2 of the Charter (which enshrines the freedom of expression). This considerations lead to conclude that no actual balancing of privacy vs freedom of expression happened.

Concluding, in statutory law the "resource privacy" conception trumps the "dignitary privacy" more than within the words of s.8 of the Charter.

2.3 The Tort claim

The third way privacy lives in the Canadian legal system is in terms of tort claims for damages. In various provinces (British Columbia, Manitoba, Saskatchewan and Newfoundland) the right to claim in case of an invasion of privacy is set out by statute.

A paradigmatic example of these provisions might be the *British Columbia Privacy Act,* which clearly states that: 'it is a tort, actionable without proof of damage, for a person, willfully and without a claim of right, to violate the privacy of another'[148]. The striking difference between this typology of claim, *actionable without proof of damage,* and its constitutional and statutory equivalent is immediately evident.

Although also in this case no clear definition of privacy is given, nonetheless the judges of the BC Court of Appeal retrieved it, defining it as "the right to be let alone, the right of a person to be free from unwarranted publicity […]. The right of an individual (or corporation) to withhold himself and his property from public scrutiny, if he so chooses[149]." This tight, brandesian definition has been widened up by a successive opinion in *Heckert v. 5470 Investments*[150] stressing the concept including the four Posser's categories[151] of actions and characterizing privacy for its attribute of being an elastic concept.

---

147 *Alberta's Personal Information Protection Act* (Alberta's PIPA) (2003).

148 Privacy Act [R.S.B.C. 1996] c. 373, s. 1.

149 *Davis v. McArthur* [1970] B.C.J. No. 664 (B.C.C.A.) at 763.

150 *Heckert v. 5470 Investments Ltd.* [2008] B.C.S.C. 1298.

In other states (e.g. Ontario), the tort is not part of the legislation but has been elaborated by the interpretative work of the judiciary, the first recognition (of 'intrusion upon seclusion') was in *Jones v. Tsige*[152]*;* judge Sharpe, drafting his opinion, building up the tort from the US one, came at the same conclusion: that of a cause of action that does not need 'proof of actual loss'.

Thus, the characterization of the tort claim about privacy breaches in Canada moves along with the concept of elasticity, in other words, a broad right, that gives rise to moral damages even in cases where no obvious losses have been suffered.

Finally, the difference of the privacy in torts is that it is strongly underpinned by a notion of inviolable personality binding it to the concept of dignitary privacy.

2.4 The future of the protection of Privacy in Canada

Considering the different characteristics of the three forms of protection for the privacy of individuals analyzed, one could hold that the concept of privacy can only develop along with its protection. Moreover, this protection, within the Canadian framework can be achieved with the synergy of three levers of this legal bulwark.

This believes underpin the 2013 report of the Standing Committee on Access to Information, Privacy and Ethics, which tries to depict the actual Canadian privacy protection system at the dawn of the Big Data age[153].

The trend followed by the commissioners in drafting the report is focused on the strengthening of the old categories elaborated within the PIPEDA.

Furthermore, the report gives great importance to a series of *privacy-enhancing methods and best practices* as the *privacy by default* setting and the *do not track* feature (on which we will focus infra).

The answer to the fragmentation and evolution of the ways the public and private organizations can penetrate the *biological core* of the individual has to be found in the achievement of an actual accountability and a real transparency. One of the examples cited in the report is the CMA (Canadian Marketing Association) Code of Ethics and Standards

---

[151] (1) Intrusion upon seclusion, (2) public disclosure of embarrassing facts, (3) false light publicity, and (4) appropriation of likeness

See William L. Prosser, "Privacy" (1960), 48 Cal. L. Rev. 383.

[152] See note 126.

[153] Standing Committee on Access to Information, Privacy and Ethics, *Privacy and Social Media in the Age of Big Data* (2013). Available at http://www.parl.gc.ca/content/hoc/Committee/411/ETHI/Reports/RP6094136/ethirp05/ethirp05-e.pdf

of Practice[154], which echoes the ten privacy principles in the PIPEDA[155] trying to give consumers control over their information and transparency to the market.

One of the most interesting topics emerging both from the academic debate and the official reports is the actual and future role of the Office of the Privacy Commissioner of Canada (OPC), having regard, in particular, to its enforcement powers[156]. These are considered too weak at the moment: under PIPEDA the Privacy Commissioner has the power to receive or initiate, investigate, and attempt to resolve complaints about any aspect of an entity's compliance with the legal provisions for data protection. However, his or her recommendations are not enforceable so the breaches of privacy legislation have to be solved through a work of negotiation and persuasion[157]. Colin Bennett has also criticized the weakness of the federal OPC for the schizophrenic coordination with provincial privacy Commissioners (Québec, BC and Alberta) who have, in fact, enforcement powers under their respective privacy laws[158].

The discussion about it compares on one hand the pros of maintaining the current model, which facilitates the flow of information between the Commissioner and private organizations promoting and incentivizing good will and self-regulation; on the other hand stand the doubts of the quality and effectiveness of the self-regulation of companies under the current model and the remedies when such self-regulatory framework fails to provide protection to individuals' privacy.

---

[154] The CMA's Code of Ethics and Standards of Practice is available at http://www.the-cma.org/regulatory/codeof-ethics

[155] (1) Accountability, (2) Identifying Purposes, (3) Consent, (4) Limiting Collection, (5) Limiting Use, Disclosure and Retention, (6) Accuracy, (7) Safeguards, (8) Openness, (9) Individual Access, (10) Challenging Compliance.

[156] Office of the Privacy Commissioner of Canada, *The Case for Reforming the Personal Information Protection and Electronic Documents Act* (2013).

available at https://www.priv.gc.ca/parl/2013/pipeda_r_201305_e.pdf

[157] See note 149.

[158] Id. p. 38.

3. The European Union Privacy legal framework today (before the GDPR)

3.1 The Fundamental Right before European Union

Privacy exists in Europe since well before the birth of EU at the threshold of the 90's of the past century. The concept and its ramification are well established in many constitutions of the member states and in the opinions of the national courts.

To fully understand the European conception of the right to privacy as a fundamental right one has to consider that Beside EU and before EU the right was firstly enshrined by the wording of the European Convention on Human Rights[159] (ECHR). The provision dedicated to the right to privacy is art.8, which states; *"[e]veryone has the right to respect for his private and family life, his home and his correspondence*"[160]. Article 8 ensures privacy rights against the intervention of governmental actors and, although it also contains lawful exceptions to its scope[161], the ECtHR has given a very broad definition of *private life*[162]. Within this broad interpretation there are the right to protection against government monitoring of employees' emails and telephone conversations to obtain evidence of improper actions at work, wiretapping phone calls without the proper checks and minimization procedures, collecting and accessing stored personal data without consent[163].

Despite the broad scope of art. 8, the technological development and the raise of the flow of data moving between individuals and organizations forced the Council of Europe to enact the *Convention for the Protection of Individuals with regard to Automatic Processing of*

---

[159] Convention for the Protection of Human Rights and Fundamental Freedoms (Nov.1950), available at http://www.echr.coe.int/Documents/Convention_ENG.pdf

[160] Id. art.8.

[161] A limitation of the privacy of the individuals can be lawfully admitted under art. 8 in the "*interests of national security, public safety or the economic wellbeing of the country, for the prevention of disorder or crime, for the protection of health or morals, or for the protection of the rights and freedoms of others*".

[162] Niemietz v. Germany, 16 Eur. Ct. H.R. 97, para. 29 (1992), available at http://hudoc.echr.coe.int/sites/eng/pages/search.aspx?i=001-57887.

[163] See *Copland v. United Kingdom*, 45 Eur. Ct. H.R. 37, paras. 43–44 (2007), available at http://hudoc.echr.coe.int/sites/eng/pages/search.aspx?i=001-79996.

Malone v. United Kingdom, 7 Eur. Ct. H.R. 14 (1984), available at http://hudoc.echr.coe.int/sites/eng/pages/search.aspx?i=001-57533.

Gaskin v. United Kingdom, 12 Eur. Ct. H.R. 36, paras. 34–37 (1989), available at http://hudoc.echr.coe.int/sites/eng/pages/search.aspx?i=001-57491.

Personal Data[164]. The recognition of the new potentiality of the computers of collecting, compiling and transferring detailed information of the individuals, pushed the member states to hold that: *"it [was] necessary to reconcile the fundamental values of the respect for privacy and the free flow of information between peoples […]"*[165].

Pursuant to this goal, the drafters of the conventions defined *personal data*[166] and *automatic processing*[167] using a language with a very broad scope. Moreover, the substantive provisions have received a broad draft too. Unlike Article 8, however, the COE Privacy Convention applies to both public and private actors[168].

In 2001 the Council of Europe added a supplementary protocol to the Convention consisting in an addition of limitations on data exportations and a commitment of the signing parties to establish independent authorities[169].

This strengthening of the privacy policy mirrors the evolution at EU level (*see infra).*

3.2 The EU Data Protection directive

In 1995 the EU passed the directive 95/46/EC[170] on the protection of individuals with regard to the processing of personal data and on the free movement of such data.

The Directive aimed to create a legal framework to govern movement of personal data across national borders within the EU and to set a baseline for the required security to

---

[164] Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data pmbl., Jan. 28, 1981, 20 I.L.M. 317 available at http://conventions.coe.int/Treaty/en/Treaties/Html/108.htm.

[165] Id. preamble

[166] Id. art. 2(a) "any information relating to an identified or identifiable individual"

[167] Id art. 2(c) the automation in whole or in part of "storage of data, carrying out of logical and/or arithmetical operations on those data, [or] their alteration, erasure, retrieval or dissemina- tion."

[168] Id. art. 3.1.

[169] Additional Protocol to the Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data regarding supervisory authori- ties and transborder data flows, Nov. 8, 2001, C.E.T.S. No. 181, available at http://conventions.coe.int/treaty/en/treaties/html/181.htm.

[170] Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data, 1995 O.J. (L 281), available at http://eurlex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:1995:281:0031:0050:EN:PDF

be provided for the storage, transmission, and processing of personal information. The text of the directive clearly refers to art. 8 of the ECHR stating that the purpose of the directive is to promote data sharing within the protective framework designed by the Convention. In other words, the twin goal of the directive was to promote the international market setting up clear standards while safeguarding a fundamental right. Under the directive member states are requested to limit the "processing of personal data", imposing restriction to this and related practices and, like in the CoE conventions the definitions of "personal data" and "data processing" are broad and comprehensive. Member states must, therefore, require that data processors collect personal data only for a specific legitimate purpose that they ensure it is accurate, and that they keep it in a form that permits identification for no longer than is necessary[171]. Consent has a key role within the directive framework and is, generally, required before processing[172].

In the directive it is also highlighted that particular categories of data must receive an enhanced protection[173]. Moreover data subject must be granted a "right to access" to personal data being processed, the individual must be put in the conditions to be aware of the purpose of the processing, the categories of data concerned, the recipients or categories of recipients to whom the data is disclosed, an intelligible form of the data undergoing processing, and knowledge of the logic involved in any automatic processing[174]. Subjects may also request that processors rectify, erase, or block data that is incomplete, inaccurate, or otherwise not in compliance with the directive, as well as notify any third parties who have received the data of this rectification, erasure, or blocking, if feasible[175].

Another element of innovation introduced via the 95/46/EC directive is the *Working Party on the Protection of Individuals with regard to the Processing of Personal Data* (also known as WP29)[176]. This Working Party is composed of representatives of each member state's data protection authorities, as well as other representatives from each member state and the European Commission[177]. The mansions of the Working Party under the directive are: examining questions of application of the directive, providing the Commission with an

---

171 Id. art. 6(1).

172 Id. art. 7 provides a list of exceptions: protection of the public interest, journalistic needs, and freedom of expression.

173 See id. art. 8(1), e.g. racial or ethnic origin, political opinions, religious or philosophical beliefs, trade union membership, health, or sexual activity.

174 Id. art. 12(a).

175 Id. art. 12(b,c).

176 Id. art. 29(1).

177 Id. art. 29(2).

opinion on the level of protection both inside and outside of the EU, advising the Commission on any proposed amendments, and giving its opinion on EU codes of conduct.

However, today the value of the data protection directive is going to become historical and doctrinal considering the advanced stage in the enactment of the General Data Protection Regulation (scheduled for 2016[178]) that will replace the directive with the power of a legal tool, such as a European regulation, which does not need implementation and transposition activities of the Member states.

3.3 The e-Privacy Directive

The initiative carried out by EU institutions of fostering the digital market, ensuring that the electronic processing of personal data would not have harmful effects on individuals' private sphere, took a further step in 2002 with the enactment of the directive on *Privacy and Electronic Communications,* also known as *e-Privacy* directive[179].

Considering the scope of the directive, the e-Privacy Directive applies to data processing conducted in connection with the provision of "publicly available electronic communications services in public communications networks"[180]. Security is the first key word of the directive; in fact the ISPs have to adopt the appropriate measures to prevent any form of data leaks and, in case of a security breach to happen, they have to duly inform the user sending a notification of what happened[181].

Furthermore, the member states are obliged, under the e-Privacy directive to enact appropriate legislative measures to ensure the confidentiality of the communications, which take electronic form[182].

---

[178] Data Protection Regulation Calendar (2012) available at http://www.europarl.europa.eu/document/activities/cont/201205/20120514ATT45081/20120514ATT45081EN.pdf

[179] Directive 2002/58/EC of the European Parliament and of the Council of 12 July 2002 concerning the processing of personal data and the protection of privacy in the electronic communications sector (Directive on privacy and electronic communications), 2002 O.J. (L 201) 37, available at http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2002:201:0037: 0047:EN:PDF

[180] Id. art. 3(1).

[181] Id. art. 4.

[182] Id. art. 5(1).

as common, in the provisions of the European legislator are also conceived limitation clauses, in the specific case they are granted for national security and criminal investigation.

National legislation must also ensure a dividing line between malicious spywares[183], software that collect information about the user without his awareness, and cookies or other legitimate forms of devices that collect information so long as consumers are provided with clear information about their purpose and are thus allowed to refuse such processing[184].

Moreover, ISPs are required to erase or anonymize data when it is no longer necessary for the purpose of the transmission[185].

With regard to location data, the provision states that ISPs cannot legitimately process them without before having made them anonymous or with an explicit opt-in consent of the user/subscriber[186].

The European people had to wait until 2009 to see the directive to be amended by the so-called *Cookie Directive*[187]. The line followed by European policymakers was in the sense of a higher security threshold for the storage and handling of personal data. In case of a security breach providers must, in addition to the notification to the user involved, inform the national data protection authorities of the states where the breach could have adversary effects[188]. The amendment also revises art. 5.3 of the e-Privacy directive, requiring the user to give consent to have their data collected in the form of cookies before third parties could store or access information in the user's device.

With the advent of the General Data Protection Regulation (GDPR) doubts have been raised about its relation with the above-mentioned directives. According to the preamble of the GDPR, this should apply to all the matters that are not subject to specific

---

[183] Id. art. 5(3).

[184] Id. pmbl., paras. 24–25.

[185] Id. art. 6(1).

as we will see *infra* this article could imply some consequences for the equivalence chosen for the two different actions of erasing/anonymizing personal data. The actual effect of this equivalence shuld be assessed comparing the different national legislations transposing the directive.

[186] Id. art. 9(1).

[187] Directive 2009/136/EC of the European Parliament and of the Council of 25 November 2009 amending Directive 2002/22/EC on universal service and users' rights relating to electronic communications networks and services, Directive 2002/58/EC concerning the processing of personal data and the protection of privacy in the electronic communications sector and Regulation (EC) No 2006/2004 on cooperation between national authorities responsible for the enforcement of consumer protection laws, 2009 O.J. (L 337) 11 available at http://eurlex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2009:337:0011:0036:en:PDF

[188] Id. art. 2(4)(c).

obligations with the same objective set out in the e-Privacy directive[189]. Thus, considered the complexity of the conflict of laws issues the same provision states that in order to clarify the relationship between GDPR and Directive 2002/58/EC, the latter directive should be amended accordingly.

[189] Proposal for a Regulation of the European Parliament and of the Council on the pro- tection of individuals with regard to the processing of personal data and on the free move- ment of such data (General Data Protection Regulation), COM (2013) 11 final (Oct. 22, 2013), available at http://www.janalbrecht.eu/fileadmin/material/Dokumente/DPR-Regulation-inofficial-consolidated-LIBE.pdf

4. The EU Privacy legal framework tomorrow (after the GDPR)

4.1 Overview of the Proposal

The GDPR has been conceived to be an adequate response to two main issues. First, the principles elaborated under the 1995 directive did not sufficiently address the rapidity of the technological development, in particular the scalar change of the online world and of the related models of business[190]; second, the harmonization was not considered the right legal tool in the field of privacy, the patchwork of national legislations implementing the '95 directive did not give the economic stakeholders the legal certainty necessary to undertake their enterprise[191].

Thus, many of the provision of the GDPR are adopted from the Data Protection directive with a greater focus on strengthening consumers' rights and promoting efficiency.

To what concerns the definitional aspects, as under the directive, the GDPR covers any processed information, concerning an identified or identifiable natural person, that forms or is intended to form part of a filing system[192], however, the GDPR expands the category of "sensitive data" including genetic data[193]. In the same way, the GDPR explicitly notes that online identifiers (emails, IP addresses, cookies etc.) are "identifiers" for the purposes of the regulation[194]. Finally, differently from the directive, the regulation drafters decided to shift the burden of responsibility from the user to the controller of data[195].

4.2 Individual control, substantive rights and transparency

Under the new regulation, implied consent will not be considered valid anymore, unlike the past regime of the directive the data controller must obtain written explicit

---

[190] Id. recital 5.

[191] European Commission, *Commission Proposes a Comprehensive Reform of Data Protection Rules to Increase Users' Control of Their Data and to Cut Costs for* Businesses, Jan. 25, 2012.

available at http://europa.eu/rapid/pressReleasesAction.do?reference=IP/12/46&format=HTML&aged=0&language=en&guiLanguage=en

[192] Compare art. 2(a) of the directive 95/46/EC and art. 4(2) GDPR.

[193] See note 198 art. 4(10).

[194] Id. recital 24.

[195] Id. recital 60.

consent for a specified purpose[196]. The theme of the consent has been specified in a stronger pro-user version; the data subject has also the right to withdraw the consent at any time[197], and, interestingly, in the first draft of the regulation consent was not considered valid where an imbalance between the data user and the controller exists[198].

Moving to one of the *new* and most controversial right: the *right to be forgotten*, outlined in the 2012 draft has been removed and substituted by the weaker but still innovative *right to erasure*[199]. Unfortunately in this circumstance it is not possible to have room for a complete analysis and comment of the new right, thus, we will outline the main features.

Data subjects will be able to require data collectors to erase the data subject's information in case of the lack of legitimate reasons to maintain it.

The change from the *right to be forgotten* to the *right to erasure* is clearly perceivable in the liability rules, under the current version the data controller can be hold liable only in case of a publication of the data that goes beyond the legitimate purposes stated in art. 6.1 GDPR, while the sketch of the right to be forgotten envisaged a full liability for the publication of data, also for the third parties[200].

As with the data protection directive, the GDPR protects the individuals from behavioral profiling[201] and data subjects retain the rights to access[202], correction[203], objection, erasure, and the right to obtain a copy of the data in an accessible format.

Beyond the reproduction of most of the categories of the data protection directive, one of the new feature characterizing the most modern approaches to the aged problem of the *notice & consent* agreements, is the introduction of a system of classification, data protection seals and marks intended to allow data subjects to quickly understand the characteristics and typology of data protection associated to each service the are going to use[204]. Transparency is the word used by the European legislator to describe the purpose of

---

196 Id. recital 25 *("Silence or inactivity should therefore not constitute consent").*

197 Id. art. 7(3).

198 in the 2012 version of the GDPR art. 7(4) ("Consent shall not provide a legal basis for the processing, where there is a significant imbalance between the position of the data subject and the controller") removed.

199 Id. art. 17(1).

200 See art 17 old version (2012) available at http://ec.europa.eu/justice/data-protection/document/review2012/com_2012_11_en.pdf

201 GDPR art. 20.

202 Id. art. 15.

203 Id. art. 16.

204 Id. recital 77.

this provision; the transparency principle helps ensuring that individuals would be able to understand the uses of their data to the extent is necessary to give consent, on the other hand prevents data collectors discriminating users according to their personal data and promotes accountability in the activity of use and maintenance of the data[205].

4.3 Accountability, control and enforcement

One of the most innovative features of the new regulation is the approach chosen by the European legislator to address the issue of accountability. In line with the academic recognition of the new categories of privacy by design/default the drafters decided to implement them in the accountability system designed for privacy in Europe. In order to incentivize the implementation of appropriate and proportionate technical and organizational measures and procedures, and to safeguard the entire lifecycle of data[206], data protection by design shall be a prerequisite for public procurement tenders[207].

The *privacy by default* stated by art. 23 requires public and private companies to ensure by default that personal data are not made accessible to an indefinite number of individuals assuring the data subject to be able to control their distribution[208].

The GDPR includes a number of provisions that shall guarantee the minimization of the data collected, retaining the information only for the necessary time and positively acting to protect them.

Moreover, in case of a security breach, beyond the notification to the data subject, the data controllers are required, under the regulation, to duly inform the supervisory authorities without undue delay[209].

The GDPR is intend to introduce a "one-stop shop" for data protection in Europe, with the implementation of the regulation business will not be required anymore to notify Data Protection Authorities in each of the countries in which they operated. Under the GDPR, a

---

[205] Id. recital 32, inspired Int'l conference of data prot & privacy comm'rs, *International Standards on the Protection of Personal Data and Privacy*, The Madrid Resolution (2009).

available at http://www.privacyconference2009.org/dpas_space/space_reserved/documentos_adoptados/common/2009_Madrid/estandares_resolucion_madrid_en.pdf

[206] Id. art. 23 *("…from collection to processing to deletion, systematically focusing on comprehensive procedural safeguards regarding the accuracy, confidentiality, integrity, physical security and deletion of personal data").*

[207] Id. art. 23(1a).

[208] Id. art. 23(2).

[209] Id. art. 31.

multinational organization will be required to contact only the Supervisory Authority of the country where its business has its main establishment[210].

The role of DPAs has been enhanced: in addition to the traditional administrative powers of notification and of forced compliance of data controllers/processors (and to bring legal action against them), they will be granted the powers of blocking the data flows to a recipient in a third country and to certify controllers according to art.39[211].

The cooperation of the different DPAs should be fostered by the new *consistency mechanism* introduced by the GDPR, intended to harmonize the application of the GDPRs provisions. The consistency mechanism can be described as the attempt of building up a hierarchical network with the European Data Protection Board[212] at the top and the national supervisory authority at the bottom. In case of application of the consistency mechanism on matters of general application the national authority has to inform the EDPB that, without undue delay shall adopt an opinion, voting with simple majority[213].

This new organ, the European Data Protection Board, takes the place of the Working Party 29 and is made up of the head of one supervisory authority for each of the Member State plus the European Data Protection Supervisor. The main duty (among a vast number) of the EDPB is to issue opinions (along with the Commission) to ensure the correct and consistent application of the GDPR[214].

Finally, in addition to the administrative remedies already granted under the DP directive, the GDPR grants the data subjects the right to act in the judiciary against a controller/processor or in response to a decision of a national Data Protection Authority[215]. To what concerns the applicable jurisdiction, the data subject can either bring legal action to the court of the place where the defendant is established or in the data subject's home jurisdiction[216]. Some mechanism of coordination are also laid down by the European legislator in order to foster the possibility for any body, organization or association which acts in the public interest to lodge a complaint if it considers that a data subject's rights under this Regulation have been infringed as a result of the processing of

---

[210] Id. art. 4(19) (defining "supervisory authority"), arts.28–29 (governing notification of the supervisory authority).

[211] Id. art.53 (1h, 1ia).

[212] Intended to replace the Working Party 29 (*see infra).*

[213] Id. Art. 58.

[214] Id. arts. 58(7), 59(1)–(2).

[215] Id. arts. 74-75.

[216] Id. art. 75(2).

personal data[217]. Coordination mechanisms for the court proceedings have been set up too[218]. Eventually, a right to a due compensation has been outlined by the words of the regulation[219].

The General Data Protection Regulation can be considered the latest stage in the development of the modern Privacy law, the regulation will pursue the process of continue integration of the members of the Union[220].

Using the words of some commentators the benefits of a system aimed to safeguard privacy as a fundamental right could only entail substantial benefits for the European people as a whole[221].

---

[217] Id. art. 73(2).

[218] Id. art. 76.

[219] Id. art. 77.

[220] M. Rotenberg and D. Jacobs, *Updating the Law of Information Privacy: The New Framework of European Union*, 36 Harvard Journal of Law and Public Policy 2 (2013).

[221] See G. Gross, *U.S. Privacy, Consumer Groups Back EU's Proposed Privacy Rules*, Com. World (2012). Available at

http://www.computerworld.com/s/article/9230931/U.S._privacy_consumer_groups_back_EU_39_s_proposed_ privacy_rules

5. The fall of (some of) the old principles about Privacy in a Big Data World

5.1 PII (Personal Identifiable Information) 2.0

The brief analysis of the legal systems conducted above tends to suggest that one of the central concepts in privacy regulation is the Personal Identifiable Information (PII)[222]. Although not every legal system clearly and explicitly defines it, the concept of PII is fundamental in two different parallel perspectives that we could define as *inner* and *outer* functions.

The two functions play their role on different fields; on the one hand, the *inner* function of PII is philosophical and it works as a sphere of existence of the individual outside of the boundaries of physical body, a trail of life. This function can be useful, and has been implicitly used[223], to outline the edges of privacy as a fundamental right.

On the other hand, the *outer* function of privacy is its work as an extraordinary instrument to define the scope and boundaries of privacy statutes and regulations[224]. All these laws share the same basic assumption: in the absence of PII no privacy harm occurs. Thus, privacy regulations are mainly focused on the collection, use and dissemination of this category of information leaving the remaining unregulated or under-regulated.

Nowadays, however, the existence itself of the PII category is posed under threat by the technological development and businesses proclivity toward big data scenarios. One of the most renowned exponents of this view, of PII as a fatally flawed concept, is Paul Ohm. In his recent article he challenged the idea of PII suggesting that privacy law should abandon its reliance on it shifting its focus on a new paradigm to regulate information privacy[225].

Despite this extremist view, other scholars tried to re-think the idea of PII without abandoning it, that is the PII 2.0, based on the idea of a binary approach to personal *identified* and *identifiable* data[226].

---

[222] P. Schwartz and D. Solove, *The PII Problem: Privacy and a New Concept of Personally Identifiable Information*, Nyu L. Rev. 1814 (2011).

[223] See note 89.

[224] Its usefulness is also related to the fact that it is a uniform instrument to investigate and study privacy regulations of different legal systems.

[225] See note 69

[226] See note 231 1817.

In the intentions of the commentators who hold this theory this should represent a step forward, that avoids both the reductionist US view of the PII and the expansionist view of EU. As a consequence, the legal protection given to the two different categories could obtain different traits.

5.2 The anonymization/re-identification dilemma

The question is: what does make data *identifiable*? The answer is that identifiable means that the individual can be identified directly or indirectly by reference to an identifier, such as a name, and identification number, a unique location etc.

This plain definition needs an integration with the operative dimension of identification or re-identification to unleash all its pitfalls. Before to analyze the power of re-identification one has to take a logical step backwards, before re-identification comes de-identification. Traditionally, de-identification has been imagined as a silver bullet, the panacea able to allow organizations to reap the benefits of analytics while preserving individuals' privacy[227]. However, all the various methods of de-identification used by organizations (anonymization, pseudonymization, encryption, key-coding, data sharing) seem to fail in guaranteeing a permanent de-identified status to the information[228].

In other words, over the past few years, analysts and computer scientists have shown that even anonymized data can be re-associated to specific individuals[229]. Thus, when thinking of de-identification, legal scholars and policymakers have to be aware of the fact than de-identified data is a temporary rather than a stable category.

This assumption shuffles the cards for all the subjects (governments and businesses) that have strongly relied on the anonymization mythology embracing it as the key factor of numerous business models, in particular in the context of clinical trials, online behavioral advertising and cloud computing.

The issue policymakers have to face is not easy to be solved.

---

[227] See, e.g., Working Party, Opinion 4/2007 on the Concept of Personal Data, Article 29 (June 20, 2007), available at http://ec.europa.eu.ezp.biblio.unitn.it/justice/policies/privacy/docs/wpdocs/2007/wp136_en.pdf

[228] E. Felten, *Does Hashing Make Data 'Anonymous'?*, Tech@FTC (2012). available at http://techatftc.wordpress.com/2012/04/22/does-hashing-make-data-anonymous;
E. Felten, *Are Pseudonyms 'Anonymous'?*, Tech@FTC (2012). available at http://techatftc.wordpress.com/2012/04/30/are-pseudonyms-anonymousi

[229] A. Narayanan and V. Shmatikov, *Myths and Fallacies of "Personally Identifiable Information*, 53 Communications of the ACM 6, 24 (2010).

Its characteristics and dimension are a crucial factor, considering the two opposing and extreme visions: on the one hand there is Ohm's conclusion that the category itself of PII has come to an end, on the other hand the specular vision that all data should be treated as PII and subjected to the pertinent regulatory framework.

While the former option has to be specified in order to assess its feasibility, the latter can be immediately excluded as a valid option, it would imply unbearable privacy management costs for economic operators and it could result in an unworkable privacy framework. Moreover, many beneficial uses of data would be consistently weakened and curtailed if privacy laws would be based on every (even remote) possibility of linking the data to an individual, the value of big data we discussed would remain latent and undisclosed.

5.3 The identifiability test – a hybrid theory

The solution of the anonymization/re-identification dilemma might be son of the same philosophical shift entailed by the entrance in the big data environment. From causality to correlation, where causality is identification and correlation is identifiability. While in case of big data analysis the algorithm to extract value from raw data is a matter for analysts, in case of the identifiability formula is up to the lawyer the task of extracting and combining the elements of the matrix.

Not an easy task, the risk matrix to apply should be built on the risk, intent, and potential consequences of re-identification, as opposed to a "identifiable/non-identifiable" stiff dichotomy. In fact this second Manichean approach is unhelpful and leads inevitably to an inefficient arms race between deidentifiers and reidentifiers, process that would result in a detrimental effect for either integrity or accuracy or value of the data together with the loss of some of its beneficial potential[230].

We believe that the test may be outlined taking the five factors suggested by Ohm in his article[231]. However we would combine them to the reconceptualization of PII operated by Solove and Schwartz[232]. Doing so we would attempt to sketch the criteria to recreate correlative boundaries to the "identifiable" category. Done that, we would further suggest to implement some of the guidelines recently elaborated by the FTC to cover the blind spots of the theory.

---

[230] See note 60 p. 258.

[231] See note 69 p. 1765.

[232] See note 231.

The five factors are the following:

Data handling techniques

Computer scientists could provide a rough relative ordering of different techniques or at least assign to them different categories of risk (e.g. low, medium, high), it is unlikely to think that it will be possible to assign a percentage of risk related to each different data handling technique, although scientists might grade favorably a database owner who implements a technique instead of another[233].

Private vs. Public Release

This second factor gives relevance of the typology of dissemination, according to the author's view a public release is far more dangerous in terms of risk of re-identification than one between trusted parties; however this factor has to be balanced against the others not resulting the public disclosure *per se* in a certain indicator of a higher degree of risk.

[233] About anonymization see V. Lakshmanan and T. Raymond, *On Disclosure Risk Analysis of Anonymized Itemsets in the Presence of Prior Knowledge*, 2 ACM Transactions on Knowledge Discovery From Data 13, 13:2 (2008). ("Among the well-known transformation techniques, anonymization is arguably the most common."). See for a comparison of methods R. Nabil and J. Wortmann, *Security-Control Methods for Statistical Databases: A Comparative Study*, 21 ACM Computing Surveys 515 (1989).

Quantity

Traditional privacy regulations focus on the quality of the information, building up different legal regimes according to the belonging or not of the information to the category. Yet in every re-identification study the scientists were helped by the size of the database. Thus policymakers could use the quantity as a factor of risk, assessing it on a case-by-case basis (unrealistic), building up database tiers based on their dimension (and to the correlated degree of re-identification risk) or introducing quantitative limits on data collection and retention[234].

Motive

In numerous contexts sensitive data are held by a small number of subjects who lack any incentive to re-identify the data set, for example the rules governing the margin of appreciation of academic research should keep into consideration that it is unlikely that this category had any interests in re-identification. On the other hand, the legislators to increase the level of risk connected should weight financial incentives. Insurance companies and health data is just an example.

Trust

The other side of motive is trust: during the age of anonymization trust was not needed, we trusted in the technology, now that the technology has fall we have to rebuild the relation beneath the processing of the data. According to the author's view we should implicitly trust academic researchers, government data miners less, and third parties advertisers not at all. After the emotive process we should finally try to draw up the conclusions into the legislation.

The joint application of the five factors could give the legislator an appreciative but useful measure of the risk involved. Moreover, the policymakers should not abandon the categories of data based on its sensitiveness, risks of re-identification of non-sensitive information are lighter on the scale of costs/benefits. If the benefits of unfettered

---

[234]One could disagree with this kind of initiatives because of their possibility of having detrimental effects on the potentiality of the processing of those data without a parallel proportionate decrease of the level of risk for such initiatives see e.g. European Union Article 29 Data Protection Working Party, Opinion 1/2008 on Data Protection Issues Relating to Search Engines, 00737/EN WP 148, at 19 (April 4, 2008), available at http://ec.europa.eu.ezp.biblio.unitn.it/justice_home/fsj/privacy/docs/wpdocs/2007/wp136_en.pdf (arguing that search engines should store queries for a maximum of six months).

information significantly outweigh the costs to privacy in a particular context, they might decide to surrender[235]. Much more often, regulators will conclude that the costs to privacy outweigh the benefits of unfettered information flow. In this case they should decide to clamp down on the information flow in targeted ways[236].

Considering the hybrid nature of the test and the legitimate doubts on how it could effectively be implemented, we might back it up, as said before, with the principles listed in a recent FTC report. This report overlays the statistical probability of re-identification, completing the effectiveness of the test with organizational commitments and downstream contractual obligations not to re-identify or to attempt to do so.

According to the FTC, the requirements that have to be fulfilled in order to escape from the scope of the – given – legal framework are that, (1) a given data set is not reasonably identifiable, (2) the company publicly commits not to re-identify it, and (3) the company requires any downstream users of the data to keep it in de-identified form[237].

Applying this second and parallel level of assessment is necessary admitting that it is virtually impossible to guarantee privacy only scrutinizing data without defining and analyzing its intended uses. In this sense the FTC policy principles are useful to close the circle of the factual test with a legal examination of the organizations' intent and commitments.

Finally, in the era of Big Data de-identification has to be seen and conceived as an important defensive measure to be taken under accountability and other data security principles, rather than a solution of the Big Data conundrum.

Our attempt of outlining a model to apply may be summed up as the attempt of measuring the technology with the tools of the legal world; aware of the difficulties of this operation we shielded the model with traditional principles, whose effectiveness is related to the (unlikely) possibility of investigation and enforcement.

---

[235] For example, Harvard's Personal Genome Project, which is sequencing the DNA of thousands of volunteers to hunt for genetic markers for disease, has essentially told its volunteers to forget about privacy. Peter Dizikes, Your DNA Is a Snitch, Salon.com, Feb. 17, 2009, available at http://www.salon.com/env/feature/2009/02/17/genetic_testing

[236] see note 69 pag.1780

[237] FTC, *Protecting Consumer Privacy in an Era of Rapid Change, Recommendations for Businesses and Policymakers* (2012).

available at https://www.ftc.gov/sites/default/files/documents/reports/federal-trade-commission-report-protecting-consumer-privacy-era-rapid-change-recommendations/120326privacyreport.pdf

5.4 Data Minimization

Although through various and different iterations and formulations, data minimization has represented a core pillar of privacy law[238]. In its essence, the data minimization principle requires organizations to collect personal data to the minimum extent possible to perform their legitimate goals. Moreover, the duration of the data retention acquires importance under this perspective in fact data have to be deleted once it is not considered useful anymore for the legitimate purpose for which they were collected the first time.

As we have noticed several times (and suggested by the wording), the Big Data business model is antithetical to the concept of data minimization; the latter incentivizes the collection of *more* data for *longer* periods of time. As we know, the "crown jewels" of big data are latent in those secondary and distant-in-time uses.

Although privacy legal systems continue to consider data minimization for what it was five or ten years ago, data minimization is simply no longer the market standard from an economic and strategic point of view. Modern organizations use to mine private, semi-public (social media), and public sources.

Thus, we could argue that, in a big data world, the principle of data minimization should be interpreted in a different way. The policymakers should require organizations to anonymize data if possible, implement standard security measures and use as the society and not only the individual as a parameter to asses the limit of the acceptable uses.

5.5 Individual control and context

As we have already briefly outlined (ch. I, par. 6.1) another fundamental pillar of privacy legal frameworks all over the world is the individual control (or consent); it is about to fall or, at least, to receive a weaker focus by the privacy legislators. As we have seen above in the U.S "notice and consent" has represented the central axis of privacy regulation for years[239]. In EU consent remains the most common way to legitimize data processing both under art. 7 of the data protection Directive and new art. 6 of the GDPR. The known issue about it is again the fact that on the one hand individuals are expected to read and understand complicated privacy disclosure clauses and give their "informed"

---

[238] See i.a. Organization for Economic Cooperation and Development, *Oecd Guidelines on the Protection of Privacy and Transborder Flows of Personal Data* p. 7-8 (1980).

[239] A shift away from notice and choice is considerably underway, as reflected in the Whitehouse Blueprint and FTC Final Report; yet, under both frameworks notice and choice remains a central principle, See note 114.

consent, on the other hand the environment in which this process takes place is increasingly complex, with data flows handled through intricate arrangements involving dense networks of platforms, including contractors, subcontractors and service providers operating globally[240].

As a commentator suggested, an analogy can be drawn between the landscape of our Big Data world and the privacy regulation in the mainframe age, with the majority of data collected by a relatively close number of entities and individuals not able to understand methods and purposes[241].

In these cases the focus has to be shifted, from the hollow shell represented by the "individual self-determination" to the duties of data protection authorities; they are provided with the necessary technological knowledge needed for running a risk assessment process, moreover they are also granted of administrative powers to enforce their decisions and recommendations[242]. Thus, the weakness of the "notice and consent" has to be noticed, however the model should not been discarded, but reshaped for Big Data and other contexts in which asymmetries in data negotiation drastically reduce users' self-determination[243].

Finally the "new" self determination system should be strengthened by increasing the three "magic" concepts all the legal scholars who tried to bring privacy law toward new scenarios are repeating as a mantra: transparency, accountability, protection-oriented architectures.

---

[240] See note 60 and note 63.

[241] See note 63, someone said that the mainframe is back again and its name is *cloud*.

[242] Id.

[243] In other words the entire ubiquitous online world, see R. Calo, *Digital Market Manipulation*, 82 Geo. Wash. L. Rev. (2014), available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=23097

6. Outlining a solution – some milestones to share the wealth

6.1 Privacy by design and Big Data

As we noticed above, one of the innovative elements of the GDPR is the fact of implementing the concept of privacy by design (PbD). Privacy by design is a prescription of building privacy directly into the design and operation, not only of the technology itself, but also of the operational systems, work processes, management structures, physical spaces and networked infrastructure[244]. The implementation of a PbD system to fight back the threat to privacy posed to the big data phenomenon can be justified by the awareness that, in this case, a technological approach could offset the negative externalities of a technological development. However, the idea of having companies to implement PbD systems cannot be described and considered as a strict technical compliance definition but it could serve as the backbone of self-regulation or legislative regulation of "responsible innovation"[245].

PbD is made up of seven principles, they are used to transform consumer privacy issues from a purely policy or compliance issue to a business element of competitiveness.

We will try to follow the path of a recent paper, which applied the PbD principles to big data organizations. In particular, a team of engineers implemented a *sensemaking* system with PbD features[246].

This case is a bright example to see the principles of PbD in action; our reconstruction of the case will move from the theoretic presentation of the principle to the practical definition of the technology (what a sense-making system is and what is it supposed to do) in order to close the circle with a synthesis of how the principles have shaped the machine.

The seven principles in their latest version are the following:

1. Proactive not Reactive, Preventative not remedial[247],

---

[244] A. Cavoukian, *Privacy by Design: The Definitive Workshop. A Foreword by Ann Cavoukian, Ph.D.*, Identity in the Information Society, 3(2), p. 247-251 (2010), available at http://www.icsd.aegean.gr/website_files/proptyxiako/78723175.pdf

[245] A. Cavoukian, *Privacy by Design in Law, Policy and Practice*, Privacy by Design (2011), available at www.ipc.on.ca

[246] A. Cavoukian and J. Jonas, *Privacy by Design in the Age of Big Data*, Privacy by design (2012), available at https://privacybydesign.ca/content/uploads/2012/06/pbd-big_data.pdf

2. Privacy as the default[248],
3. Privacy embedded into design[249],
4. Full functionality[250],
5. End-to-End Lifecycle Protection[251],
6. Visibility and transparency[252],
7. Respect for user privacy[253].

"Sensemaking" relates to a new class of technology designed to help organizations make better sense of their diverse observational space[254]. Sensemaking systems will handle extremely large datasets, generated by an increasing number of different sources. These sensemaking techniques integrate new transactions (observations) with previous transactions using this context-accumulation process to improve understanding about what is happening right now[255].

In late 2008 Jeff Jonas and his team embarked in a journey to embed privacy-enhancing properties within a "sensemaking-style" system. The team tried to weigh

---

[247] See note 253, p 249: "It does not wait for risks to materialize, nor does it offer remedies for resolving infractions once they have occurred—it aims to prevent them from occurring".

[248] Id., p. 250: "No action is required on the part of the individual to protect their privacy—it is built into the system, by default".

[249] Id.: "Privacy is integral to the system, without diminishing functionality".

[250] Id.: "It avoids the pretense of false dichotomies, such as privacy vs. security, demonstrating that it is possible to have both".

[251] Id.: "Privacy, having been embedded into the system prior to the first element of information being collected, extends throughout the entire lifecycle of the data involved, from start to finish".

[252] Id., "Its component parts and operations remain visible and transparent, to users and providers alike".

[253] Id. "Keep it user-centric—focused on the individual."

[254] See note 255 p. 4.

[255] Id., p.5. More generally in the literature, sensemaking refers to a set of meta-theoretical assumptions that lead explicitly to an overall approach to framing questions, gathering data, and conducting analyses for arriving at substantive theory. This approach has been under development, primarily through the communications research of Brenda Dervin, since 1972, but has since been guided by other disciplines. Sensemaking's core assumption is that of discontinuity. There are gaps between entities which include other people, artefacts, systems, or institutions. Information seeking is associated with these 'cognitive gaps' in our understanding. Filling the cognitive gaps in our understanding is much like asking for street directions in a foreign country.

performance consequences, default settings and which PbD features would be so deeply wired within the system that they simply cannot be disabled without disabling the whole.

The following part of the paragraph will focus on the feature that Jonas and his team addressed in this new generation sensemaking system, the features are strictly technical and difficult to understand for a non-engineer or IT expert, however we believe it is important to insert them in this work, trying to give a practical example of what PbD means and how it works considered its fundamental role in the new privacy regulations frameworks (GDPR above all).

1. *Full Attribution*: every observation (record) needs to know from where it came and when. There cannot be merge/purge data survivorship processing whereby some observations or fields are discarded[256].
2. *Data Tethering*: adds, changes and deletes occurring in systems of record must be accounted for, in real time, in sub-seconds[257].
3. *Analytics on anonymized Data*: the ability to perform advanced analytics (including some fuzzy matching) over cryptographically altered data means organizations can anonymize more data before information sharing[258].
4. *Tamper-Resistant Audit Logs:* every user search should be logged in a tamper-resistant manner — even the database administrator should not be able to alter the evidence contained in this audit log[259].
5. *False Negative Favoring Methods:* the capability to more strongly favor false negatives is of critical importance in systems that could be used to affect someone's civil liberties.
6. *Self-Correcting False Positives*: with every new data point presented, prior assertions are re-evaluated to ensure they are still correct, and if no longer correct, these earlier assertions can often be repaired — in real time[260].

---

[256] J. Jonas, *Source Attribution, Don't Leave Home without It*, J.Jonas Blog (2006), available at: http://jeffjonas.typepad.com/jeff_jonas/2006/10/source_attribut.html

[257] Id., *Data Tethering: Managing the Echo*, J.Jonas Blog (2006), available at: http://jeffjonas.typepad.com/jeff_jonas/2006/09/data_tethering_.html

[258] Id., *To Anonymize or Not Anonymize, that is the Question*, J.Jonas Blog (2007), available at: http://jeffjonas.typepad.com/jeff_jonas/2007/02/to_anonymize_or.html

[259] J. Jonas, *Immutable Audit Logs (IAL's)*, J.Jonas Blog (2006).
available at: http://jeffjonas.typepad.com/jeff_jonas/2006/02/immutable_audit.html

7. *Information Transfer Accounting:* every secondary transfer of data, whether to human eyeball or a tertiary system, can be recorded to allow stakeholders (e.g., data holders or the consumers themselves) to understand how their data is flowing[261].

We share the authors' believe that building in privacy-enhancing elements into technology can help to minimize the privacy harm. This would build a higher degree of confidence in the stakeholders, contributing to a faster growth of the big data industry, aiding an overall dissemination of its beneficial effects (considering the privacy harms as negative effects).

6.2 Transparency

"Sunlight is said to be the best of disinfectants", these words belong to Louis Brandeis, co-father of the right to privacy[262].

Some legal scholars firmly believe that if the existence and uses of databases were visible entering the public sphere, organizations would be more likely to avoid unethical, socially unacceptable and discriminatory uses of data outflowing from big data analytics[263]. Moreover, if organizations were forced to disclose their line of reasoning in data processing, by law, contract or best practices, they might avoid unethical uses of data pertaining to certain populations (e.g. children, seniors etc) or data of categories that could be subject of discrimination.

Transparency, in the same way of confidentiality, fosters trust being able to hold others accountable[264]. Furthermore, transparency inherently includes tension between

---

[260] J. Jonas, *Self-Correcting False Positives/negatives: Exonerate the Innocent*, J Jonas Blog (2012). Available at: http://jeffjonas.typepad.com/jeff_jonas/2012/05/self-correcting-false-positivesnegatives-exonerate-the-innocent.html

[261] Id., *Out-Bound Record-Level Accountability in Information Sharing Systems*, J Jonas Blog (2007). Available at: http://jeffjonas.typepad.com/jeff_jonas/2007/12/out-bound-recor.html

[262] L. Brandeis, *What Publicity Can Do*, Harper's Weekly (1913). available online at http://c0403731.cdn.cloudfiles.rackspacecloud.com/collection/papers/1910/1913_12_20_What_Publicity_Ca.pdf

[263] See note 60, p. 270.

[264] N. Richards and J. King, *Big Data Ethics*, 49 Wake Forest L. Rev. 393 (2014).

openness and secrecy and this tension can generate paradoxes for both entities and individuals: transparency of sensitive corporate or government secrets could result in harmful outcomes for important interests, such as trade secrets and national security. On the other and opposite hand, too little transparency could result in a lack of trust with a connected "chilling" effect on the economic and social interactions between subjects.

The important role of transparency has been heightened with the advent of big data analysis practices[265]; the power of the secondary uses, so typical of the big data industry, has been targeted by data brokers as a source of wealth. However, the category of data brokers has been recently attacked for not meeting many of the FIPs, especially those relating to transparency.

Related to this issue, in December 2012 the FTC launched a privacy probe over the data broker industry's collection and use of consumer data[266]. In another recent report stated that the lack of data broker transparency regarding the source of data and use has the main effect of exacerbates an "aura of secrecy surrounding the industry[267].

Finally, enhanced transparency it is a fundamental requirement for a world that wants to face the challenged posed by big data analysis, it will deter unethical, sensitive data use and relieve concerns about the risk of inaccurate inferences.

6.3 Big Data due process: a peculiar model

In a 2014 work Kate Crawford and Jason Schultz outlined a new model to fight back the threats posed by the big data leviathan, in particular they focused their legal analysis on the harms of a discriminatory use of the data, performed by private subjects, such as employers, insurance companies etc. Their idea, interesting in the essence, is to apply the principle of the procedural due process to data[268].

---

[265] A. Watters, *What Does Privacy Mean in an Age of Big Data?*, O'Reilly (2011), available at http://strata.oreilly.com/2011/11/privacy-big-data-transparency.html (documenting an interview with author Terence Craig on the importance of transparency in the age of big data).

[266] K. Bachman, *FTC Launches Probe of Data Broker Privacy Practices*, Adweek (2012), available at: http://www.adweek.com/news/technology/ftc-launches-probe-data-broker-privacy-practices-146041

[267] A. Tanner, Senate Report Blasts Data Brokers for Continued Secrecy, Forbes (2013). available at http://www.forbes.com/sites /adamtanner/2013/12/19/senate-report-blasts-data-brokers-for-continued-secrecy/

[268] K. Crawford and J. Schultz, *Big Data and Due Process: Toward a Framework to Redress Predictive Privacy Harms*, 55 B.C. L. Rev 93 (2014). The authors took inspiration from a 2010 article, in which Danielle Citron stressed the idea of applying due process to automated systems. See also D. Citron, *Technological Due Process*, 85

The main issue this approach wants to tackle are the lack of any meaningful notice related to predictive mechanisms. Secondly, also in case the notice occurred, providers are unlikely to give any form of justification of their reasoning or to provide evidence of the reasons behind.

Thus, the structure of the (Big) data due process ought to be formed by three main principles, which need a further specifications and implementations. They are the following: (1) notice, (2) opportunity for a hearing, (3) impartial adjudicator and judicial review.

Notice

This principle takes two different shapes depending on its application *before* or *after* the adjudication (or query) performed by the service provider. Before the adjudication it entails the right of an individual of being aware of the fact that its data could be included or is going to be included or used in a predictive adjudication, in the moment of the consent. The second, and for us the most interesting, aspect of the notice principle applied to data is timely placed after the adjudication and should entail the notice – to those who were or are being affected – of the issues predicted, data considered and methodology employed. Moreover, the notice should also provide for a mechanism to access the audit trail or record that were created in the predictive process[269].

Opportunity for a hearing

Once the notice is available, the second step is to provide the individuals who received it the legal toolkit to make use of it. The subject shall have the chance to challenge the fairness of the predictive process employed. The belief of the authors is that establishing the possibility to be heard would entail the examination of the evidence used, including the data used and the algorithm applied. The role should be given to a trusted third party whose role would be to act as a neutral data arbiter, used to routinely examine big data providers whose adjudications may give rise to predictive privacy harms and litigations.

Wash. U. L. Rev. 1249, 1256 (2008). "Danielle Citron examines the use of automated systems in governmental administrative proceedings, the risks they pose to deprivations of liberty and property, and how a reinvigorated approach to due process could help mitigate and address".

[269] C. Dwork and D. Mulligan, *It's Not Privacy, and It's Not Fair*, 66 Stan. L. Rev. Online 35, 36-38 (2013), suggesting that bias testing could be beneficial to decrease and hamper privacy harms. See also Consultative Comm. of the Convention for the Prot. of Individuals with Regard to Automatic Processing of Pers. Data [ETS No. 108], Propositions of Modernisation, COUNCIL OF EUR. 4-5 (Dec. 18, 2012),

The importance of the presence of a neutral data arbiter is increased by the fact that, often, big data excludes any user participation

Impartial adjudicator and judicial review

Another famous myth about Big Data is that the outputs of this new approach to data analysis are generally supposed to be free of bias or at least closer to the objective truth (that does not exist) than other forms of knowledge[270]. Procedural due data process serves as a remedy in these cases, it may represent a valuable framework for ensuring a greater fairness within the predictive analysis system.

A neutral data arbiter, position played in the European system by the national data protection authorities, could file a complaint and investigate in case of sufficient allegations of bias, searching for financial interests that could result in an unfair adjudication.

The due process model is aware of the challenges posed by Big Data, nevertheless it offers a wide range of legal weapons to counter each challenge with a common purpose: ensuring protections that could be defined as both fair and feasible for all the subjects at stake.

6.4 Toward a holistic approach to privacy

The paragraphs above provide a quick perspective of how difficult could be to imagine a unitary solution for the big data issue(s).

It is a matter of fact, and a historical modification of the threat to a well-defined object: individual privacy. Scholars, policymakers and (surprisingly) entrepreneurs are elaborating methodologies to fight against privacy breaches on a daily basis and their achievement are, as we tried to underline *supra*, quite remarkable.

However, the price to pay in order to retrieve a coherent system of protection is incredibly high for the civil lawyer. This price is the admission that the modern privacy protection system cannot be unitized anymore (if ever this logic operation has been possible). The scalar jump of the data analytics requires the modern rule-makers and interpreters to take a step backwards and leave the walls of a bypassed citadel to engage a war on different fronts. We are aware of the sense of weakness and confusion such a multi-layered conception of protection might instill, however it is a sacrifice we suggest undertaking with a purpose-oriented mindset.

[270] See note 277.

The today protection of privacy is a complex interlacing set of different safeguards: general principles, regulatory provisions, administrative decisions and actions of the privacy authorities, best practices, and privacy embedded within the technology and more[271].

This awareness is the logic background of a new holistic approach to privacy. We will be asked to pay a high toll and it might result in forms of schizophrenic behavior of the different actors, on the other hand the lack of aesthetic features of the holistic approach is repaid by a stronger multi level and tridimensional sphere of protection for Privacy.

Finally, the big data golem (in its privacy harm specification) will need the single layers to be coherent to them and coordinated. This is actually the path undertaken by the EU, aiming to the protection of the individual and setting up all the necessary tools to protect it. On the other hand, the idea of the zero-sum risk depicted in the context of privacy by design[272] should be taken as a model to assess the economic impact of the privacy protection system, trying to balance it against the costs borne by the private and public organizations.

[271] We have to consider the fact that in the different legal systems the balance (or unbalance) of the different means of protection could differ due to their legal traditions, systems, conventions.

[272] See note 253.

III. Big Data and IP law – a cost-benefit analysis

1. The disclosure dilemma

One of the main questions still unsolved about the development of a Big Data world is whether there is a valid reason why big data analysis has not delivered yet the main body of the huge innovations predicted by the commentators[273].

According to technology experts, the answer to this unpleasant question lies within the folds of the challenges of data reuse[274]. The reasons behind this difficulty in building up a system based on the free movement of data to boost the positive effects of their reuse are multiple. Beside the substantial impediments that prevent data from being effectively reused, one set of challenges is purely technical and deal with the format in which data is often recorded and published. Researchers have often to work with data recorded in a wide variety of formats so they may encounter difficulties in aggregating data from multiple sources[275]. Hopefully this and similar problems will be overcome in time, good signals in this direction are already recognizable; for instance, the U.S National Institute of Standards and Technology (NIST) assembled a working group dedicated to Big Data with the aim of developing a common set of definitions, taxonomies and reference architectures[276].

The second typology of barriers to a general and wider widespread of data for reuse is less evident but more challenging; the point is that data is often deeply infused with the subjective judgments of those who collect and organize it[277]. The point is, as recently remarked by an expert in the field who has highlighted the phenomenon, that hidden biases in the analysis stages present considerable risks and their role is as relevant as that of the numbers themselves in the big-data equation[278].

All these embedded judgments consist in a problem for data reuse because data re-users may not be able to know or find the exact trail of the prior actors, so later interpretations may directly depend upon multilevel inferences that are statistically

---

[273] M. Mattioli, *Disclosing Big Data*, 99 Minn. L. Rev. 535 (2014).

[274] C. Borgman, *The Conundrum of Sharing Research Data*, 63 J. Am. Soc'y for Info. Sci. & Tech. 1059, 1059-60 (2012).

[275] M. Madison, *Commons at the Intersection of Peer Productions, Citizen Science, and Big Data: Galaxy Zoo*, Governing Knowledge Commons 209 (2014).

[276] NIST, Big Data, Nat'l Inst. Standards & Tech (2013). Available at http://bigdatawg.nist.gov

[277] I. Lawal, *Ensuring the Integrity, Accessibility, and Stewardship of Research Data in the Digital Age*, 10 portal: Libraries and the Academy 365 (Johns Hopkins University Press 2010).

[278] K. Crawford, *The Hidden Biases of Big Data*, Harv. Bus. Rev. Blog (2013).

problematic. Likewise, other commentators hold that it is often difficult to interpret and make use of the data when you don't understand how the data were generated[279].

Nevertheless, in some context a focus to data and data practices disclosure already exists but with different purposes; in some academic research settings as leading scientific and economic journals, authors are required to submit information about their data sources and detailed descriptions of the specific techniques the used to prepare the data for the study[280].

However, and beyond that, the market reason that hampers a full disclosure of data collected by those actors who are in the position to do it (providers of search engines, mobile, health devices, public utilities), mostly collect that data more as a byproduct than as an actual direct source of business, the lack of market for such abstract information has the consequence of giving little impetus to disclosure[281]. Thus, as we already said, big data represents a largely speculative value that for its own nature resides far downstream from the commercial exchanges that take place between data producers and their customers.

Finally, the other face of the lack of any affirmative economic incentive to disclose is the number of solid disincentives the economic and institutional actors have to deal with. In particular, privacy regulations might impede the conveying of information about their anonymization practices of institutions that collect and transmit PII records. Competition concerns, closing the overview, might discourage disclosure of data preparation methodologies. Likewise, it is unlikely that big data giants would desire a disclosure of information that might be used against them to claim, for instance, a weakness in their methods (resulting in a low quality data).

The fast track to innovation promised by the big data experts is meeting a number of technical, commercial, and epistemological roadblocks that are significantly slowing and limiting the data's potential for a future reuse.

---

[279] L. Peer, *Mind the Gap in Data Reuse: Sharing Data Is Necessary But Not Sufficient for Future Reuse*, London Sch. Econ. & Poli. Sci. (2014).

Available at http://blogs.lse.ac.uk/impactofsocialsciences/2014/03/28/mind-the-gap-in-data-reuse

[280] Editorial, Social Software, 4 Nature Methods 189 (2007), available

at http://www.nature.com.ezp.biblio.unitn.it/nmeth/journal/v4/n3/full/nmeth0307-189

(requiring authors to submit all algorithms, software and related data)

[281] D. Boyd and K. Crawford, *Six Provocations for Big Data* (2010).

Available at http://ssrn.com/abstract=1926431

(discussing what little **data** Twitter releases to researchers and the problems resulting from lack of **disclosure** of storage methods).

Aware that some of these impairments are difficult avoidable or represent the price to pay on the altar of our fundamental rights, we will try to study how and if intellectual property law might serve as a useful way to alleviate the grasp of all these factors on the pace of innovation.

2. The influence of Intellectual Property law upon disclosure

One of the known goals of IP law is to spur innovation by encouraging technological dissemination[282], in the field of big data, for a variety of reasons we tried to outline IP law is not meeting this goal. In order o assess this statement it is first necessary to study how current IP constructs may apply to Big Data practices.

Hopefully the novelty of the phenomenon will not impede the attempt of framing it within the existing intellectual property system. In fact, debates about intellectual property and algorithms, software and database are a longstanding challenge considered on their own, in this case they sum up so the question is whether the single categories have to be considered separately or joint in a new form.

The following sections will try to scrutinize the role of trade secret, patent and (briefly) copyright in order to lay the groundwork for further speculations.

2.1 Trade secret

In the US the Uniform Trade Secrets Act (UTSA) provides the definition for trade secret, which has been adopted by most of states.

It is defined as "information" that is (i) valuable, and (ii) reasonably protected[283]. This definition can be described as highly expansive, covering both technical and non-technical information, including methods, know-how and ideas too[284]. Most notably, information does not need to be absolutely secret to receive dignity under the UTSA, it must only be subject of reasonable efforts to prevent disclosure[285].

The remedies for a misappropriation of trade secrets may range from monetary damages to injunctive relieve. On the other side of the Atlantic there is no uniformity of

---

[282] M. Lemley, *The Surprising Virtues of Treating Trade Secrets As IP Rights*, 61 Stan. L. Rev. 311, 332 (2008).
(Beside that a second purpose (of IP law) - some argue the main one - is to ensure that the public receives the benefit of those inventions).

[283] Unif. Trade Secrets Act § 1(4), 14 U.L.A 538.
for another definition see also Restatement (Third) of Unfair Competition § 39 (1995) (A trade secret is any information that can be used in the operation of a business or other enterprise and that is sufficiently valuable and secret to afford an actual or potential economic advantage over others").

[284] V. Chiappetta, *Myth, Chameleon or Intellectual Property Olympian? A Normative Framework Supporting Trade Secret Law*, 8 Geo. Mason L. Rev. 69, 76 (1999).

[285] See note 286.

the European legislation, as noted in a 2013 study made by the EU commission in order to prepare a proposal for a directive aimed to align existing laws against the misappropriation of trade secrets across the EU. The report clearly underlines the lack of a uniform definition of trade secret in Europe; the consequence is that the criteria to be met in order to acquire the trade secret protection are different from member state to member state[286]. Nonetheless the absence of a common definition, some common traits can be recognized: (i) the information is technical or commercial and inherent to the business activity, (ii) the information is secret, meaning that they are not part of the public domain and they are not easily accessible, (iii) the information has an economic value that gives the owner a competitive advantage, (iv) the information is subjected to technical measure that prevent the disclosure, or at least, make it more difficult[287].

In particular, information-based processes that are non-readily perceived by consumers are particularly well suited for trade secret protection. For this reason, it represented the heart of the academic debate about software and intellectual property in the 90s, the issue at that time is similar to the one we are facing in these work but with some important differences.

At that time, leading intellectual property scholars argued that the use of trade secret, with the implied cost of discouraging the disclosure of source code and related practices, would slow the pace of software innovation. Robert G. Bone, for instance, highlighted the significant costs of trade secret, it would lead to wasteful duplicative efforts among software engineers working at different firms[288]. Pamela Samuelson cautioned that

---

[286] Study on trade secrets and confidential business information in the internal market (2013) available in french at http://ec.europa.eu/internal_market/iprenforcement/docs/trade-secrets/130711_executive-summary_fr.pdf

("Des définitions légales spécifiques sont uniquement proposées par la législation suédoise sur les secrets d'affaires, les Codes de la Propriété Intellectuelle italiens et portugais ainsi que les législations applicables à la concurrence déloyale de la Bulgarie, de la République Tchèque, de la Grèce, de la Pologne et de la République Slovaque. En Hongrie et en Lituanie, la définition légale est prévue dans leur Code Civil respectif. En Slovénie, une définition figure dans la Loi relative aux Sociétés. Lorsqu'aucune définition formelle des secrets d'affaires n'est prévue, la notion est circonscrite par la jurisprudence : ceci est le cas en Autriche, en Belgique, à Chypre, au Danemark, en Estonie, en Finlande, en France, en Allemagne, aux Pays-Bas, en République d'Irlande, en Lettonie, au Luxembourg, à Malte, en Roumanie, en Espagne et au Royaume-Uni").

[287] Id.

[288] R. Bone, *A New Look at Trade Secret Law: Doctrine in Search of Justification*, 86 Cal. L. Rev. 241, 266-67 (1998).

secrets are expensive to keep[289]. All summed up, from all these and similar insights, legal scholars warned that such a high-rate widespread of trade secrecy would reduce the cumulative rate of innovation in the software industry.

Some scholars at that time identified the relief valve in the fact that software methods can be sometimes reverse engineered, in this direction has been argued that trade secrecy was not a complete bar to the dissemination of the know-how embedded in software technologies because reverse engineering is permitted by law and not so difficult to perform on object code[290]. Furthermore, Mark Lamley identified a second potential benefit of trade secrecy that could encourage investors in spend less money in building physical barriers to maintain their secrets[291].

For Big Data practices the things are not so straightforward, like algorithms, many big data practices likely fit within trade secret law's expansive definition of "information"[292]. Moreover, secrecy over this kind of information may be even easier to keep than the same level of secrecy over a software product; in fact it has been underlined that, unlike software, big data practices in most of the cases cannot be reverse-engineered.

As a result, the assertion of the commentators that trade secrecy may sometimes promote disclosure of software methods seem to be inapplicable to big data practices.

2.2 Patent

The patent option could theoretically push the developers of some Big Data practices toward public disclosure. Assuming for a moment that the broad category of "Big Data practices" could comprehend some inventions under the US patent act[293], the applicant would be forced to fill in the application with a detailed written description of the invention claimed.

In return, as we all know, patentees receive a far more robust form of protection than trade secret holder could ever enjoy: the possibility of preventing any unauthorized

---

[289] P. Samuelson, *A Manifesto Concerning the Legal Protection of Computer Programs*, 94 Colum. L. Rev. 2308, 2329 (1994).

[290] J. Reichman, *Computer Programs as Applied Scientific Know-How: Implications of Copyright Protection for Commercialized University Research*, 42 Vand. L. Rev. 639, 701(1989).

[291] See note 286

[292] "Trade secret' means information, including a formula, pattern, compilation, program, device, method, technique, or process."

[293] The same rationale could be applied, changing what has to be, to any western patent act or regulation.

use, manufacture, sale, or importation of their innovation for twenty years[294]. Despite the protection advantages, the point is that it is not sure that patent protection could fit Big Data practices: first of all, patent protection extends to a narrower set of processes and methods than trade secrecy. Thus, algorithms that amount to abstract ideas, for instance, do not meet the threshold eligibility requirement to be grant the patent protection[295]. Moreover, only processes that are novel, non-obvious and useful may be eligible[296]. While the utility bar seems not to represent an obstacle for big data practices, it is still unclear whether they are sufficiently novel and non-obvious to receive the patent protection.

Furthermore, even if patent protection would be available to provide legal shelter to information protecting methods, trade secret seems to be still preferable according to economists. In a landmark article, two economists identified a double pattern of situations in which trade secret trumps patent: when patent protection seems too costly relative to the value of the invention, or when patent protection would provide a reward substantially lower than the value of the invention[297].

The direct consequence is that the perceived cost of obtaining the patent and the perceived value of secrecy could lead a Big Data actor not to submit the patent application, even in case it would be legally feasible.

2.3 Copyright

This overview of the relation between Big Data and intellectual property would be not entirely complete without a quick look at copyright. One of the main differences with patent law is that copyright does not grant exclusivity in processes or methods.

Nonetheless, in some cases copyright may protect the products of such practices. The argument supporting this statement is that originality, main element of the copyright, has been found in data estimates, classifications and in compilations arranged through methods that strongly rely on upon subjective human judgment[298].

---

[294] Patent Act, 35 U.S.C. § 112 (2012).

[295] See Alice Corp. v. CLS Bank Int'l, 134 S. Ct. 2347 (2014) (holding that adding a computer to perform a set of functions that are otherwise abstract ideas does not confer patentability).

[296] Patent Act, 35 U.S.C. §§102-103.

[297] D. Friedman, Some Economics of Trade Secret Law, 5 J. Econ. Persp. 61-64 (1991).
for instance, if an invention could be easily kept secret for a period of time longer than it would take other inventors to come up with the idea on their own.

[298] See, e.g., CCC Info. Servs., Inc. v. Maclean Hunter Mkt. Reports, Inc., 44 F.3d 61, 67 (2d Cir. 1994) (stating that individual estimates of used car prices published by plaintiff were "original creations" for

However, copyright has the slight disadvantage that it does not require the authors of compilations to disclose their methods of compilation. That means that only if copyrightability is challenged in a court of law this information could be disclosed. Thus, its seems to be a poor candidate to promote the disclosure of big data practices.

purposes of copyright); Am. Dental Ass'n v. Delta Dental Plans Ass'n, 126 F.3d 977, 979 (7th Cir. 1997) (holding short numerical codes copyrightable subject matter). The Copyright Act explicitly protects compilations "selected, coordinated, or arranged in such a way that the resulting work as a whole constitutes an original work of authorship." Copyright Act, 17 U.S.C. § 101 (2012).

3. Industry practices

The disclosure dilemma would not be completely understood without surveying the characteristics of Big Data practices. Two themes in particular have emerged from investigations on big data practices: these practices are highly subjective and almost impossible to discover through reverse engineering. Moreover, a number of disincentives to disclose, both economic and legal, work to push toward secrecy. Those findings vary, however, according to the different type of Big Data practice analyzed. The presentation of these differences will follow the four primary Big Data practices: filtering non-relevant data (i.e. "noise") from large datasets, identifying and correcting errors based on estimates or guesses, "masking" data in order to preserve anonymity, classifying data.

3.1 Searching the Haystacks

The task of locating useful information within a large corpus of data represents the ultimate search for the needle in the haystack. The problem to face is that online sources used by big data providers (social networks, online forums, etc.) span a huge array of topics and often are full of "noise" in the form of spam.

As a result, data obtained from these sources need to be sifted out and sorted before being used.

An increasing number of technology startups boast special expertise in sifting data, some of them assemble information on a vast number of topics, other companies focus on a single topic[299].

The matter of implied subjectivity in judgments needed to sift out huge amasses of data can be clearly highlighted by an anecdote provided by TrueLens, a Boston-based firm, which operates with direct advertising. Supposing an airline to decide the launch of two new routes from Boston and San Francisco to Denver. The airline has a list of its past customers, but it does not know which of these customers are likely to be interested in the Boston-Denver route versus the San Francisco-Denver route. This is where Big Data sifting steps in. By analyzing publicly available information about the airline's customers (e.g., information that customers opted to share publicly on their social media profiles, publicly posted photos, check-ins and comments, etc.), the company is able to identify

[299] i.a. DataSift provides its customers with specialized streams of data culled from the hundreds of millions of daily posts made to social networks. In contrast Treato automatically collects the massive amount of the content patients generate online.

which of the airline's past customers are more likely to be interested in one particular route over the other.

Thus, significant human judgment goes into assembling this data, at TrueLens, some researchers have a hunch, for example, that customers most interested in the airline's new route are those who live in major cities and who also enjoy skiing.

3.2 Cleansing

The raw datasets that data analysts work with often contain errors in different forms. As a result, unprecedented volumes of data imply an unprecedented number of errors. A second source of errors is the automatic and indiscriminate operation of information gathering that is the hallmark of the big data method. Third and even more subtly, some data errors manifest when error-free databases are merged.

In practice, identifying and correcting such errors is as much an exercise in aesthetics as statistics[300].

Because data cleaning is often highly subjective, different practitioners could easily reach different final products.

To better understand how data cleaning works data expert and economist from a prominent social network offered a helpful but hypothetical example.

Suppose a Big Data analyst working for an online business wishes to collect data on how long visitors stay on her employer's website. When the analyst collects relevant data from the company's web server, she finds that most visitors appear to stay on the website for 2-5 minutes. Some of the data doesn't make sense, however: the server reports many visits lasting "0 minutes" in length, some visits lasting several days in length, and a few inscrutable results such as "infinity" and "not a number".

Faced with these anomalous results, the analyst might first try to find the sources of the errors. She may guess, for instance, that the records of visits lasting "0 minutes" were generated by automated software agents known as "bots". Users who walked away from their computers without closing their web browsers, meanwhile, probably generated the visits apparently lasting for days. Lastly, she surmises that a bug in the web server's software caused the reports of "infinity" and "not a number".

[300] See note 277, p.561.

3.3 Masking and Suppression

Either to comply with legal regulations, or for self-regulatory best practice, many big data producers use to obfuscate or mask personal identifying information contained in the raw data they work with.

In fact, even in absence of a legal mandate, market forces have pushed some big data producers to mask personal data. As the same way as the previous categories of practices analyzed, data masking represents a mix of science and art, a product often infused with subjective judgments.

Entering the essence of the practice, the easiest way to anonymize a dataset is to strip it of information that could be used to identify individuals: names, addresses, zip codes, phone numbers. This approach is definitely one of the most successful in ensuring individual privacy but presents the drawback of destroying a high percentage of the value of the data.

A secondary and less destructive option is to systematically replace personally identifiable information with dummy values. In such a way it is possible to identify the same individuals over time[301]. Furthermore, data masking sometimes involves techniques far more complicated that just replacing names. Experts at CancerLinQ can offer an example, a project organized by the American Society of Clinical Oncologists in 2012[302]. CancerLinQ aggregates clinical information from hospitals around the country relating to cancer treatment. Such information includes, for instance, lab tests and doctors' notes. The system then culls this data and correlates the successfulness of treatments with patient characteristics in order to provide treatment suggestions.

Experts working on CancerLinQ turned to a software firm that specializes in de-identifying patient data. This software allows the users of the system to prioritize the preservation of key information as well as permitted permutations, such as shifting all treatment dates equally to preserve a longitudinal record of the length of a particular patient's treatment without reporting actual dates of treatment. Here too, subjectivity plays a central role, at every step of the way there are a lot of subjective questions and answers. For example, the person using the software must be able to say how much they trust the recipient of the data or whether they think data might be publicly exposed.

---

[301] R. White, Web-Scale Pharmacovigilance: Listening to Signals from the Crowd, 20 J. Am. Med. Informatics Ass'n 404 (2013). Available at http://jamia.bmj.com.ezp.biblio.unitn.it/content/20/3/404.full.pdf

[302] CancerLinQ, Am. Soc'y of Clinical Oncology (2014), available at http://www.asco.org/quality-guidelines/cancerlinq

3.4 Classifying

The fourth and last analyzed technique of manipulating data is classification, used to alter data prior to publication. The idea behind classification is the same of some optical illusion that show, for example a cube emerging from a surface or entering the surface depending on the elements on which the sight focuses. In the same way, the picture drawn by Big Data is often in the eyes of the beholder. It is the climax of personal subjective judgment; classifications and taxonomies show the degree of personal perception that Big Data practitioners impose upon the data they work with.

Classification is fundamental in Big Data applications that cull linguistic data for insights; an increasing number of startups focus in this practice of so-called "sentiment analysis"[303]. Although elaborated software might perform some of such classifications on its own, human judgment is almost always required to make accurate and useful categories out of linguistic data.

The classification of data requires often and appreciation for context that only a human can judge. An example may be useful; in 2011 Dr. Stephens of HS University gathered and presented scores of online Twitter posts in a map of the U.S. that identifies where hateful speech is most prevalent[304].

In carrying out this project, Dr. Stephens considered the fact that identifying "hate" is more difficult than simply searching for certain words. In fact, she realized and understood that, depending on context, some derogatory terms can take on a positive or negative connotation. To address this problem, Dr. Stephens had to ask her assistants to manually review each post, in order to remove those not derogatory in nature, and then classify the speech in the posts that remained.

Thus, the final processed dataset reflects subjective classifications that were made by Dr. Stephens and her research team. This brief anecdote shows how classifying data to facilitate analysis is a key big data practice, and like data sifting, it appears to sometimes entirely rely upon subjective human judgments.

---

[303] S. Baker, *The Numerati* (Houghton Mifflin Harcourt 2008), pp 43-65 (discussing the practice of dividing consumers into "buckets").

[304] M. Stephens, *FAQ: Geography of Hate*, Floating Sheep (2013). Available at http://www.floatingsheep.org/2013/05/hatemap.html

4. Intellectual property implications and suggestions

The foregoing paragraph will question whether intellectual property is a suitable candidate to encourage the disclosure of Big Data practices. Anticipating the response one can only say that the answer is, for the most part, negative.

Big Data practices seem not to fit neatly within the current intellectual property legal scenario and related paradigms. At the same time, the trait of these practices of being not self-disclosing lends them well to trade secret or to mere nondisclosure status.

These elements lead to sketch the features of a possible IP right to encourage disclosure of such practices.

4.1 Why not patent law

The main reason why patent law does not seem to be a meaningful candidate to encourage the disclosure of big data practices is the following: the practices above analyzed in particular and big data practices in general appear either unlikely to meet patent law's threshold eligibility requirements, or potentially eligible but unlikely to provide a meaningful scope of protection.

The first impairment for a useful application of patent law to these practices is the high degree of subjectivity we discussed before, the direct consequence of subjectivity from a patent law point of view is the ineligibility for the lack of sufficient definiteness to claim them.

However, in other cases is possible to define a big data practice with more precision: cleaning and data masking might be likely objectively anchored. Although this typology of practices could be underpinned by an objective structure, other barriers to patent protection may nevertheless stand.

For instance, a failure in showing sufficient non-obviousness or novelty could lead to a direct rejection; likewise methods of preparing data could be denied patent protection when consisting in mere abstract ideas.

Eventually, even with patent protection available, big data providers may nevertheless prefer to follow the path of nondisclosure. The scenario is clear, the lack of economic interest in patents is also related to the fast-paced dimension of the big data industry, the economic value of these practices is relatively short-lived and, as a result, not worth the time and trouble of obtaining patent protection.

Beyond the lack of legal incentives to disclosure, there are a number of disincentives. The first of them is the system of privacy regulations; it is the case of data

masking and suppression practices that, once disclosed, would make re-identification definitely easier.

Likewise, data producers feel like disclosing their methods and operations for dealing with data may highlight flaws in their methodologies or weakness in their underlying data. Finally, and as a result, intellectual property framework is not the only factor in the non-disclosure of big data practices.

4.2 Why not Copyright?

Copyright seems to be inadequate too, offering a surprisingly thin protection for corpora of big data.

As already mentioned before, copyright law can protect original expression found in compilations of data. The category of data sifting and manipulation listed and described above clearly meet the originality bar as forms of selection[305]. In other words, human creativity plays a key role in the selection and arrangements of data as we tried to underline. However, from a practical point of view, such protection is unlikely to be an effective means to curtail unwanted reproduction of methods. The reason is that the final product could, in theory suffer partial copy, with the impossibility of claiming the originality of the sifting process because the single or multiple data copied do not represent or consist in an original method. The metaphor of the theft of the single tiles of a mosaic is an effective way to describe the phenomenon.

Classification too seems to meet the copyright's originality requirement; however, the American case law shows a stark contrast between circuit courts.

The copyrightability of classifications that reflect subjective judgments has been hold by the 2007 case of American Dental Association v. Delta Dental Plans Association, the Seventh Circuit held that individual six-digit codes for dental procedures were copyrightable works of authorship that met Copyright's originality threshold[306]. In the same case, Judge Eastbrook found that the plaintiff's placement of related procedures in similar

---

[305] In the words of Copyright Act, 17 U.S.C. § 101 (2012): "compilations selected, coordinated, or arranged in such a way that the resulting work as a whole constitutes an original work of authorship". An important limitation on this form of copyright, however, is that it "extends only to the material contributed by the author of such work ... and does not imply any exclusive right in the preexisting material", Id. § 103.

[306] Am. Dental Ass'n v. Delta Dental Plans Ass'n, 126 F.3d 977, 979 (7th Cir. 1997) (holding the short numerical codes to be copyrightable subject matter).

numerical series was an expression of judgment that met copyright originality threshold. The key sentence from Judge Eastbrook was that: "originality is a creative endeavor"[307].
On the other hand, other circuits have explicitly refused to provide copyright protection in similar cases.
In a 2004 case of Southco, Inc v. Kanebridge Corporation, judge Alito of the Third circuit explained that offering copyright protection would contrast with the fundamental and longstanding tenet that protection may not extend to words of short phrases[308].
The risk that this extension of copyright protection could potentially lead anyone who uses a given number to become an infringer pushed the overwhelming majority of judgments on the topic to opt for the denial of such protection. As a result, big data providers cannot rely on copyright to prevent unwanted copying of data classification.

4.3 A third way to disclosure

Assumed the lack of disclosure in the big data practices market as a flaw to correct for the above-listed reasons, one should be aware that the problem is not inherently an "intellectual property problem". Rather, IP is relevant in the discourse for what concerns issues of technological disclosure, and in this sense can play an important role and give an essential contribution[309].
The purpose of this paragraph is to assess to what extent an intellectual property solution would be helpful.
Given the insufficiency of the traditional IP tool to tackle the issue effectively, Mattioli sketches the main traits of a new *sui generis* right named, without a great creative effort, *dataright*[310].
Although this dataright shares with the EU database right the feature of having been created ad hoc, nonetheless the scope of application and typology of protection granted

---

[307] Id.

[308] Southco, Inc. v. Kanebridge Corp., 390 F.3d 276, 285-87 (3d Cir. 2004) (en banc). An additional basis for denying protection was that, unlike the dental classifications in Delta Dental, the screw fastener numbers were arbitrarily selected and as a result, "totally unoriginal". Id. at 289 (Becker, J., concurring).

[309] B. Frischmann, *Infrastructure: The Social Value of Shared Resources,* Oxford University Press 263 (2012). As Brett Frischmann has noted, "Intellectual property laws are a prominent but by no means exclusive means of addressing the supply-side problem where free riding is a concern and appropriating benefits through market exchange of the intellectual resource or some derivative product is relevant to investment decisions".

[310] See note 277, p.578

differ consistently, according to the different purpose (the recognition of the investment made in one case, the trigger to disclosure in the other).
This dataright would be available to applicants who disclose clear and complete descriptions of their methodology of preparation and collection of data alongside the final product, the data shaped following those methods.
The new legal construct will be formed upon three main elements that characterize nearly all the forms of intellectual property rights:

1. scope of the subject matter covered by the right,
2. exclusive rights conferred to publishers of this subject matter,
3. the set of acquisition rules and requirements upon which exclusivity is conditioned.

To what concerns the first point, the subject matter covered by the dataright might consist in any data that has been collected or manipulated according to one or more methods not readily apparent to a person of ordinary skill in the art[311].
Turning to exclusivity, the structure of the right requires for dataright holders the entitlement to sue unauthorized users of their data or methods for injunctive relief for a limited period of time. Similarly to the patent's rationale we might wish a limited exclusive entitlement aimed to balance data producers' desire to prevent downstream use against the public's interest in having widespread access to data. To reach this optimal balance the dataright holder might prevent, for instance, unauthorized uses, not reproduction or distribution of the descriptions of the subject matter that entail the underlying data too, performed by third parties. No matter the contractual obligations not to do so[312].

[311] This is taken from patent law, which invokes the "person of ordinary skill" to resolve issues pertaining to initial protection.

[312] In the U.S. intellectual property framework Data producers have long relied upon contracts to curtail unwanted copying. U.S. Copyright Office, *Report on Legal Protection for Databases*, 22 (1997), available at http://www.copyright.gov/reports/db4.pdf. ("For many database producers, contracts provide a major source of protection, either complementing copyright law or picking up the thread where it falls short"). This method of "self-help" in the data publishing industry may prevent some unwanted copying, but publishers have long lamented that contracts alone are far weaker than intellectual property protection because they avail only against licensees and not against unlicensed downstream copyists.

Turning to acquisition requirements, the right would be granted to those big data producers who disclose all data collection and relevant organization and arrangement practices, for each piece of data the seek to protect under the dataright[313].

The difference with the disclosure required for the patent application is noticeable considering that the subject matter they protect (data) would be different from the subject matter they disclose (methods)[314].

The range of effectiveness of this hypothetical right is still to determine, at the current state of art, data publishers have demonstrated that could need a sui generis protection to acquire a greater control over the downstream uses of their data[315].

The economic theory rationale behind this phenomenon explains that data producers will prefer dataright protection over trade secrecy each time they value the exclusivity over the downstream uses of their data more that exclusivity in their practices. This leads to consider that such a legal construct would be ineffective in all those situations in which privacy or strong commercial incentives push toward secrecy.

Moreover, another set of challenges a sui generis right may face are political, the U.S. congress is used to consider and not approve bills designed to provide sui generis protection for electronic databases[316]. And beyond data specific issues, sui generis rights are in general problematic, As Mark Janis and Stephen Smiths have hold, specialized forms of intellectual property protection designed around specific technologies tend to be inherently inflexible and might result in a reduction of consistency and predictability of the entire system of intellectual property as a whole[317].

To shield the dataright from the critics that would consider intellectual property rights upon data as a way to undermine the competitive ethos on which market economies depend, one could argue that a dataright would not entitle a data provider to impede

---

313 Interestingly this plan is part of the notion of semi patent, see G. Parchomovsky, M. Mattioli, *Partial Patents*, 111 Colum. L. Rev. 207 (2011).

314 It is known in the academic field that the reach of intellectual protection is never perfectly coextensive with the degree of disclosure required. See id p. 208.

315 See The Consumer and Investor Access to Information Act of 1999: *Hearing on H.R. 1858 Before the Subcomm. on Telecomms., Trade, & Consumer Prot.* of the H. Comm. on Commerce, 106th Cong. 67-68 (1999).

316 Differently from EU that established its database right under the directive 96/9/EC available at http://eur-lex.europa.eu/lexUriServ/LexUriServ.do?uri=CELEX:31996L0009:EN:HTML

317 M.D. Janis, S. Smith, *Technological Change and the Design of Plant Variety Protection Regimes,* 82 Chi.-Kent L. Rev. 1557, 1560 (2007).

the copy or distribution of data[318]. Instead, it would be mainly aimed to prevent unauthorized use of data.

Setting political challenges aside for a moment, the most significant risk for the success and effectiveness of dataright involves the selective nondisclosure preformed by data producers. In other words, data producers could provide inadequate, vague or incomplete descriptions of their methods of operation in order to achieve protection. A solution might consist in the doctrine of inequitable conduct that provides, in patents, that applicants who made misrepresentations of the reality during the application process may have their patents invalidated[319].

The last challenge is economic, the new benefits promised by dataright are likely to bring alongside new costs that could consist in a disincentive for disclosure. These costs can be distinguished in application costs and litigation costs. Considering the latters, the hypothetical litigation of dataright may consist of two elements: whether a purported use of data may constitute infringement, and whether a given disclosure would be sufficient to receive exclusivity in exchange.

The approach to these challenges should be profiled considering the expertise of courts in dealing with "old" IP categories, arguing the fact that dataright maintains the same structure and intellectual approach; and the work of policymakers in providing expertise and competency to assess whether the level of disclosure fit the purposes of the legislation.

Eventually, aware that there may be no room for intellectual property based solutions to the big data disclosure problem, including data in the closed group of objects of protection could provide economic and social benefits able to outweigh the significant costs that this plan would entail.

---

[318] To what concerns the arguments against IP rights in data, SEE J.H. Reichman , P. Samuelson, *Intellectual Property Rights in Data?,* 50 Vand. L. Rev. 51, 164 (1997).

[319] U.S. Patent & Trademark Office, Manual of Patent Examining Procedure § 2016 (9th ed. 2014).

Conclusions

"What's past is prologue," wrote the Bard. This sentence perfectly fits our idea of Big Data, nothing more than the algorithmic version of the same concept.

The effects of big data, on our everyday live, are large on a practical level, as an effect of the technology applied to find new solutions to old issues.

But it is just the start. The reason is that old certainties and consolidated categories are being challenged. Our worldview built upon causality is being threatened by the impressive predictive power of correlations.

One could hold that big data seems to represent the paramount fulfillment of the promises of the "information society" giving materiality to its name. One of the few certainties we have for the future developments of the phenomenon is that the amounts of data will not stop to grow, as well as the computational power needed to crunch them.

As we tried to underline in the first chapter, Big Data is giving new room to our ability to do more, faster and better, unleashing new added value and generating new winners and losers.

One should always recall that Big Data's predictions are not set in stone, nor they are necessarily completely true in all their parts; they are "just" likely outcomes meaning that a modification chance exists and remains an available tool to interact with.

The risks implied within the massive collection of data coming from a daily growing number of sources are mainly related to threat/protection of the private sphere of individuals. Our comparative analysis of three fundamental legal systems has highlighted that policymakers are generally aware of the need for the classical privacy and data protection legislations to be rethought. From the draft of the Consumers' privacy Bill of Rights in the U.S., to the upgrade of the regulatory and enforcement powers of state data protection authorities in Canada, to the draft of the General Data Protection Regulation in the EU context. The common political (and legal) goal is to create an environment free from threats to give a playfield to economic actors and consumers to interact within the boundaries of the phenomenon, unleashing the highly promised wave of innovation of Big Data analysis.

From a legal and academic perspective, the revolution for analytics brought by big data has weakened the well-established privacy legal principles. Our task has been that of highlighting the weak spots of the old categories of thought and filling the gaps of a cracked, flawed but not destroyed bulwark.

From the importance of an up-to-dated definition of PII to the degree of weakness of an anonymization system on which the confidentiality of our health data rely. The solution of the Big Data privacy conundrum we tried to outline is based on a multi-layered system of protection. Keeping the old principles adapting their traits to the new technological categories of tools able to interact with our individuality, providing sufficient legal instruments to the entities entitled to protect it and giving relevance to new modality of protection for our privacy that may partially shift the battleground from the legal to the technology world (privacy by design and privacy by default above all). This approach to the new privacy issues has been defined as holistic, in the above sense of a general sensibility toward privacy matters.

From privacy dilemmas to intellectual property economic incentives, the big data phenomenon and its multidimensional relevance have become a reality for the legal world too.

The attempt of fitting the old IP structures with the new practices of a Big Data world has showed all its practical unfeasibility. However, it has not been in vane, reasoning on the peculiar features of the phenomenon one can sketch the elements needed to be implemented in a hypothetical intellectual property right in order to have it to work with big data.

In a fast and movable landscape such as the Big Data practices one, we have realized that the classic IP rights might be sided by a newborn, light and dynamic dataright that would require further studies and debate to overcome the seeable impairment related to the nature of the practices and of the phenomenon.

Summarizing the core of this work, we tried to briefly outline how a technological and scientific shift is reverberating its effect on the real world, with subtle waves of innovation and danger, we tried to depict the current state of art of a western legal world aware of the issue and that placed it on a top position in its legislative agenda.

Furthermore, we hope that the analysis of the different protection techniques from different legal traditions and coming from different legal backgrounds we listed and highlighted should become source of insights and inspirations.

Finally, we must consider that big data analysis and its related legal framework are just tools for a better and more balanced playfield; in the same way predictive analysis is not the answer of all our current and future questions and legal tools and regulations will not give all the answers to the fears of invasion of our personal globe of information.

Thinking at this last statement from an image; if Henry Ford would have been able to query a big data algorithm for what his customers wanted, it would have answered "a faster horse". In the legal world after Big Data explosion we shall not give up on the human "ingenuity" (compared to the strict rationality of the algorithm) as source of progress.

Concluding about big data and the world, we shall remember that the information we are able to collect and process will be always only a tiny fraction of the information that exists, the idea is the same of that of the shadows of Plato's cave, vague simulacrum of the real world.

In the same way and with the same approach the scholar and the legislator have to face the consequent challenges as those arising from a tool that does not offer ultimate answers, with the same pragmatism and intellectual agility, which characterize the world of technological innovation.

Bibliography

Anderson C., *Free: The Future of a Radical Price* (Hyperion 2009)

BBC, 'Everyone "to Be Research Patient", Says David Cameron' *BBC UK Politics* (5 December 2011) <http://www.bbc.com/news/uk-16026827> accessed 2 October 2015

BBC, 'Telefonica Hopes "Big Data" Arm Will Revive Fortunes' *BBC Technology* (9 October 2012) <http://www.bbc.com/news/technology-19882647> accessed 2 October 2015

Bachman K., 'FTC Launches Probe of Data Broker Privacy Practices' [2012] Adweek

Bagley A. and Brown J., 'Limited Consumer Privacy Protections against the Layers of Big Data' [2014] 31 Santa Clara Computer & High Tech. L.J. 483

Baker S., *The Numerati* (Houghton Mifflin Harcourt 2008)

Barrass R. and Wasser L., 'Seclusion Intrusion: A Common Law Tort for Invasion of Privacy' [2012] McMillan LLP

Berman J., 'Principles of Big Data Preparing, Sharing, and Analyzing Complex Information' [2013] Amsterdam: Elsevier

Beyer M., 'Gartner Says Solving "Big Data" Challenge Involves More Than Just Managing Volumes of Data' (Gartner 2011)

Bone R., 'A New Look at Trade Secret Law: Doctrine in Search of Justification' [1998] 86 Cal. L. Rev. 241, 266-67

Borgman C., 'The Conundrum of Sharing Research Data' [2012] 63 J. Am. Soc'y for Info. Sci. & Tech. 1059, 1059-60

Boyd D. and Crawford K, 'Six Provocations for Big Data' *SSRN eLibrary. Retrieved from http://ssrn.com/paper=1926431 on.*

Brandeis L., 'What Publicity Can Do' [1913] Harper's Weekly

Brito J. et al., 'Crowdsourcing Government Transparency' [2008] 9 Sci. & Tech. L. Rev. 119

Brown B., et al., 'Are You Ready for the Era of "Big Data?' [2011] McKinsey Quarterly

Brynjolfsson E., et al., 'Strength in Numbers: How Does Data-Driven Decision-Making Affect Firm Performance?' (*A51*, 2011) <http://www.a51.nl/storage/pdf/SSRN_id1819486.pdf>

Calo R., 'Digital Market Manipulation' [2014] 82 Geo. Wash. L. Rev. 995

Cavoukian A., 'Privacy by Design: The Definitive Workshop. A Foreword by Ann Cavoukian, Ph.D.' [2010] Identity in the Information Society

Cavoukian A., 'Privacy by Design in Law, Policy and Practice' [2011] Privacy by Design

Cavoukian A. and Jonas J., 'Privacy by Design in the Age of Big Data' (2012) Privacy by design

Chiappetta V., 'Myth, Chameleon or Intellectual Property Olympian? A Normative Framework Supporting Trade Secret Law' (1999) 8 Geo. Mason L. Rev. 69

Chopra A., 'Modeling a Green Energy Challenge after a Blue Button', (2011) The White House Office of Science and Technology Policy

Chopra A., 'Green Button: Providing Consumers with Acces to Their Energy Data.' [2012] Office of Science and Technology Policy Blog

Chopra A., et al., 'Blue Button' Provides Access to Downloadable Personal Health Data' (White house Blog 2010)

Citron D., 'Technological Due Process' (2008) 85 Wash. U. L. Rev. 1249

Clarke R, 'An Evaluation of Privacy Impact Assessment Guidance Documents' (2011) 1 Int'l data privacy law 111

Cohen J., 'What Privacy Is For' (2013) 126 Harv. L. Rev. 1904

Conley A., et al., 'Sustaining Privacy and Open Justice in the Transition to Online Court Records: A Multidisciplinary Inquiry' (2012) 71 Md. L. Rev. 772

Cox J. and Cline K., 'Parsing the Demographic: The Challenge of Balancing Online Behavioral Advertising and Consumer Privacy Considerations' (2012) 15 J. Internet L. Rev. 3

Crawford K., 'The Hidden Biases of Big Data' (2013) Harv. Bus. Rev. Blog

Crawford K. and Schultz J., 'Big Data and Due Process: Toward a Framework to Redress Predictive Privacy Harms' (2014) 55 B.C. L. Rev 93

Cukier K., 'The Mood of the Market' (2012) The Economist Online

Davenport T., Barth P. and Bean R., 'How Big Data Is Different' (2012) Sloan Review 43-46

De Vries W., 'Protecting Privacy in the Digital Age' (2003) 18 Berkeley Technology Law Journal 283

Dekkers M., et al, 'Measuring European Public Sector Information Resources' (2006) Mepsir

Diebold F.X., 'Advances in Economics and Econometrics: "Big Data" Dynamic Factor Models for Macroeconomic Measurement and Forecasting: A Discussion of the Papers by Lucrezia Reichlin and by Mark W. Watson' (2003) 37 Ec. Soc. Mon. 46

Diebold F.X. 'A Personal Perspective on the Origin(s) and Development of "Big Data": Phenomenon, the Term, and the Discipline' (2012) 13 Pier Working Paper 3

Duhigg C., 'How Companies Learn Your Secrets' (2012) N.Y. Times Magazine

Dwork C. and Mulligan D, 'It's Not Privacy, and It's Not Fair' (2013) 66 Stan. L. Rev. Online 35

Eur. Data Prot. Supervisor, 'Privacy and Competitiveness in the Age of Big Data: The Interplay between Data Protection, Competition Law and Consumer Protection in the Digital Economy.' (2014) Preliminary opinion

European Commission, 'Commission Proposes a Comprehensive Reform of Data Protection Rules to Increase Users' Control of Their Data and to Cut Costs for Businesses' (2012)

Executive Office of the President, 'Big Data: Seizing Opportunities, Preserving Values' (The White House reports 2014)

'Executive Order 13642: Making Open and Machine Readable the New Default for Government Information' (US President Barack Obama 2013)

FTC, 'Protecting Consumer Privacy in an Era of Rapid Change, Recommendations for Businesses and Policymakers' (2012)

Felten E., 'Are Pseudonyms "Anonymous"?' (2012) Tech@FTC Blog Online

Felten E., 'Does Hashing Make Data "Anonymous"?' (2012) Tech@FTC Blog Online

Frieden J., Price C and Murray L, 'Putting the Genie Back in the Bottle: Leveraging Private Enforcement to Improve Internet Privacy' (2011) 37 Wm. Mitchell L. Rev. 1671

Friedman D., 'Some Economics of Trade Secret Law' (1991) 5 J. Econ. Persp. 6

Golder S. and Macy M., 'Diurnal and Seasonal Mood Vary with Work, Sleep, and Daylength Across Diverse Cultures' (2011) 333 Science 1

Grey J., 'Towards a Genealogy of Open Data.', *General Conference of the European Consortium for Political Research.* (2014)

Gross G., 'U.S. Privacy, Consumer Groups Back EU's Proposed Privacy Rules' (2012) Com. World 1

Grunes A. and Stucke M., 'No Mistake about It: The Important Role of Antitrust in the Era of Big Data.' (2015) 4 Am. Bar Ass. 22

Hughes R., 'Two Concepts of Privacy' (2015) 31 Com. Law & Sec. Law Rev. 4

Int'l conference of data prot & privacy comm'rs, 'International Standards on the Protection of Personal Data and Privacy', *The Madrid Resolution* (2009)

Jamieson A., 'Smart Meters Could Be "Spy in the Home"' (2009) *The Telegraph*

Janssen K., 'Towards a European Framework for the Re-Use of Public Sector Information: A Long and Winding Road' (2003) 11 International Journal of Law and Information Technology 184

Janssen K., 'The Influence of the PSI Directive on Open Government Data: An Overview of Recent Developments' (2011) 28 Government Information Quarterly 446

Jonas J., 'Data Tethering: Managing the Echo' (2006) J.Jonas Blog

Jonas J., 'Immutable Audit Logs (IAL's)' (2006) J.Jonas Blog

Jonas J., 'Source Attribution, Don't Leave Home without It' (2006) J.Jonas Blog

Jonas J., 'Out-Bound Record-Level Accountability in Information Sharing Systems' (2007) J.Jonas Blog

Jonas J., 'To Anonymize or Not Anonymize, That Is the Question' (2007) J.Jonas Blog

Jonas J., 'Self-Correcting False Positives/negatives: Exonerate the Innocent' (2012) J Jonas Blog

Kalil T., 'Big Data Is a Big Deal' (*The White House*, 2012) <https://www.whitehouse.gov/blog/2012/03/29/big-data-big-deal>

Kaplan R. and Norton D., 'Strategy Maps: Converting Intangible Assets into Tangible Outcomes.' (2004) Harvard Business Review Press

Lafave W., 'Search and Seizure: A Treatise On The Fourth Amendment' (2011) 78 Mich. Law Rev. 451

Lakshmanan V. and Raymond T., 'On Disclosure Risk Analysis of Anonymized Itemsets in the Presence of Prior Knowledge' (2008) 2 ACM Transactions on Knowledge Discovery From Data 13

Laney D., '3D Data Management: Controlling Data Volume, Velocity and Variety' (2001) Meta Group Report

Laney D., 'To Facebook You're Worth $80.95' (2012) Blog WSJ Online

Lawal I., 'Ensuring the Integrity, Accessibility, and Stewardship of Research Data in the Digital Age' (2010) 10 Nat. Ac. Press 365

'Leadership Under Challenge: Information Technology R&D in a Competitive World An Assessment of the Federal Networking and Information Technology R&D Program' (President's Council of Advisors on Science and Technology 2007) <https://www.nsf.gov/geo/geo-data-policies/pcast-nit-final.pdf>

Lemley M., 'The Surprising Virtues of Treating Trade Secrets As IP Rights' (2008) 61 Stan. L. Rev. 311

Leone M. and Robotti N., 'Enrico Fermi E La Presunta Scoperta Dei Transuranici' (2003) Atti Del XXIII Congresso Nazionale Di Storia Della Fisica E Dell'astronomia

Levin A. and Nicholson M.J., 'Privacy Law in the United States, the EU and Canada: The Allure of the Middle Ground' (2005) 5 Ottaw. Univ. Law & Tech. J. 357

Lewi J. and Cukier K., 'Data, Data Everywhere' (2010) 3 <http://www.economist.com/node/15557443> accessed 2 October 2015

Madison M., 'Commons at the Intersection of Peer Productions, Citizen Science, and Big Data: Galaxy Zoo' (2014) Governing Knowledge Commons 209

Mantelero A., 'The Future of Consumer Data Protection in the E.U. Re-Thinking the "notice and Consent" Paradigm in the New Era of Predictive Analytics.' (2014) Computer Law & Science Review, 652.

Mattioli M., 'Disclosing Big Data' (2014) 99 Minn. L. Rev. 535

Mayer-Schonberger V. and Cukier K., *Big Data: A Revolution That Will Transform How We Live, Work, and Think* (Eamon Dolan/Houghton Mifflin Harcourt 2013)

Maykuth A., 'Utilities' Smart Meters Save Money, but Erode Privacy' (2009) The Philadelphia Inquirer

McDonald A. and Cranor L., 'The Cost of Reading Privacy Policies' (2008) 4 I/S: J.L. & Pol'y 540

Michel J., et al., 'Quantitative Analysis of Culture Using Millions of Digitized Books' (2011) Science Online

Moringiello J. and Reynolds W., 'From Lord Coke to Internet Privacy: The Past, Present, and Future of the Law of Electronic Contracting' (2013) 72 Md. L. Rev. 452

Moussavi M., 'A Model for Quality of Schooling', (2010) *Proceedings of Aaai Artificial Intelligence for Development.*

NIST, 'Big Data' (2013) Nat'l Inst. Standards & Tech 24

Nabil R. and Wortmann J., 'Security-Control Methods for Statistical Databases: A Comparative Study' [1989] 21 ACM Computing Surveys 515

Narayanan A., 'Robust De-Anonymization of Large Sparse Datasets' (2008) Iee Symp. On Security & Privacy 111

Narayanan A. and Shmatikov V., 'Myths and Fallacies of "Personally Identifiable Information' (2010) 53 Communications of the ACM 6

Nazareth R. and Leite J., 'Stock Trading in U.S. Falls to Lowest Level since 2008' (2012) Bloomberg Online

Newman N., 'Search, Antitrust and the Economics of the Control of User Data' (2014) 31 Yale J. On Reg. 401

Office of the Privacy Commissioner of Canada, 'The Case for Reforming the Personal Information Protection and Electronic Documents Act' (2013)

Ohm P., 'Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization' (2010) 57 Ucla L. Rev. 1701

'Opinion 2/2010 on Online Behavioral Advertising , at 5, WP 171' (Article 29 Working Party 2010)

Organization for Economic Cooperation and Development, 'Oecd Guidelines on the Protection of Privacy and Transborder Flows of Personal Data'

Organization for Economic Cooperation and Development, 'Thirty Years After The OECD Privacy Guidelines p.17'

Owano N., 'Engineers Unleash Car-Seat Identifier That Reads Your Rear End' (2011) <http://phys.org/news/2011-12-unleash-car-seat-rear.html>

Peer L., 'Mind the Gap in Data Reuse: Sharing Data Is Necessary But Not Sufficient for Future Reuse' (2014) London Sch. Econ. & Poli. Sci. Online

Podesta J. and Pritzker P., 'Big Data: Seizing Opportunities, Preserving Values' (1 May 2014) <https://www.whitehouse.gov/sites/default/files/docs/big_data_privacy_report_may_1_2014.pdf> accessed 2 October 2015

Prosser W.L., 'Privacy' (1960) 48 California Law Review 383

Quinn E., 'Smart Metering and Privacy: Existing Law and Competing Policies' (Colorado Public Utility Commission 2009)

Reichman J., 'Computer Programs as Applied Scientific Know-How: Implications of Copyright Protection for Commercialized University Research' (1989) 42 Vand. L. Rev. 639

Richards N. and King J., 'Big Data Ethics' (2014) 49 Wake Forest L. Rev. 393

Rotenberg M., 'Foreword: Privacy and Secrecy after September 11' (2002) 86 Minn. Law Rev. 1115

Rotenberg M. and Jacobs D., 'Updating the Law of Information Privacy: The New Framework of European Union' (2013) 36 Harvard Journal of Law and Public Policy 2

Rubinstein I.S., 'Big Data: The End of Privacy or a New Beginning?' (2013) 3 International Data Privacy Law 74

Samuelson P., 'A Manifesto Concerning the Legal Protection of Computer Programs' (1994) 94 Colum. L. Rev. 2308

Schwartz P. and Solove D., 'The PII Problem: Privacy and a New Concept of Personally Identifiable Information' (2011) 1 Nyu L. Rev. 1814

Sitaram A. and Huberman B., 'Predicting the Future with Social Media', *Proceedings of the 2010 Ieee/wic/acm International Conference on Web Intelligence* (2010)

'Smart Privacy for the Smart Grid: Embedding Privacy into the Design of Electricity Conservation' (Information and privacy commissioner of Ontario & future of privacy forum 2009)

Solove D., 'Access and Aggregation: Public Records, Privacy and the Constitution' (2002) 86 Minn. L. Rev. 1137, 1160-64

Solove D., 'Conceptualizing Privacy' (2002) 90 California Law Review 10

Solove D., 'A Taxonomy of Privacy' (2006) 154 University of Pennsylvania Law Review 477

Solove D., 'Privacy Self-Management and the Consent Dilemma' (2013) 26 Harvard Law Review 1880

Standing Committee on Access to Information, Privacy and Ethics, 'Privacy and Social Media in the Age of Big Data' (2013)

Stanley J., 'Eight Problems with "Big Data' (2012) Aclu Blog

Steadman I., 'Big Data, Language and the Death of the Theorist (Wired UK)' (25 January 2013) <http://www.wired.co.uk/news/archive/2013-01/25/big-data-end-of-theory>

Stephens M., 'FAQ: Geography of Hate, Floating Sheep'

Story L. and Helft M., 'Google Buys DoubleClick for $3.1 Billion' (2007) N.Y. Times Magazine

Stuart D., 'The Unfortunate Dilution of Section 8 Protection: Some Teeth Remain' (1999) 25 Queens L. J. 65

Taipale K., 'Data Mining and Domestic Security: Connecting the Dots to Make Sense of Data' (2003) V The Columbia Science and Technology Review

Tanner A., 'Senate Report Blasts Data Brokers for Continued Secrecy' (2013) Forbes

Tavani H., 'The Consent Process in Medical Research Involving DNA Databanks: Some Ethical Implications and Challenges' (2010) 40(2) ACM SIGCAS Computers and Society 11

Tene O. and Polonetsky S., 'Big Data for All: Privacy and User Control in the Age of Analytics' (2013) 11 Northw. J. of Tech. and Intel' Prop. 239.

The Boston Consulting Group, 'The Value of Our Digital Identity' (2012) <http://www.libertyglobal.com/PDF/public-policy/The-Value-of-Our-Digital-Identity.pdf> accessed 2 October 2015

'The Open Society: Governments Are Letting in the Light' (2010) The Economist

The White House, 'Consumer Data Privacy in a Networked World: A Frame Work for Protecting Privacy and Promoting Innovation in the Global Digital Economy'

(2012) <https://www.whitehouse.gov/sites/default/files/privacy-final.pdf> accessed 3 October 2015

Thomas J., 'Project Gutenberg Digital Library Seeks To Spur Literacy' (2007) U.S. Department of State, Bureau of International Information Programs

Toole J., 'Quantifying Crime Waves', *Proceedings of Aaai Artificial Intelligence for Development* (2010)

Tucker D., 'Big Mistakes Regarding Big Data' (2015) American Bar Association ABA: Antitrust Source.

Van Brakel R. and De Hert P., ') Policing, Surveillance and Law in a Pre-Crime Society: Understanding the Consequences of Technology Based Strategies' (2011) 20 J. Police Stud. 163

Warren S.D .and Brandeis L.D., 'The Right to Privacy' (1890) 4 Harvard Law Review 193

Watters A., 'What Does Privacy Mean in an Age of Big Data?' (2011) O'Reilly

Weiss S.M. and Indurkhya N., *Predictive Data Mining: A Practical Guide (The Morgan Kaufmann Series in Data Management Systems)* (1st edn, Morgan Kaufmann Publishers In 1997)

Westin A. et al. *Privacy and Freedom* (Scribner 1967)

White House Report, 'Consumer Data Privacy in a Networked World: A Framework for Protecting Privacy and Promoting Innovation in the Global Digital Economy' (2012)

White R., 'Web-Scale Pharmacovigilance: Listening to Signals from the Crowd' (2013) 20 J. Am. Med. Informatics Ass'n 404

Wired and Anderson C., 'The End of Theory: The Data Deluge Makes the Scientific Method Obsolete' (23 June 2008) <http://archive.wired.com/science/discoveries/magazine/16-07/pb_theory> accessed 2 October 2015

Zittrain J., 'Meme Patrol: When Something Online Is Free, You 'Re Not the Customer, You 'Re the Product' (2012) The Future of The Internet

List of Authorities

Authors Guild, Inc v HathiTrust [2014] 755 F3d 87 (2d Cir)

Authors Guild, Inc v Google, Inc [2011] 770 FSupp2d 666 (SDNY)

Davis v. McArthur [1970] B.C.J. No. 664 (B.C.C.A.) at 763

Dental Ass'n v. Delta Dental Plans Ass'n, 126 F.3d 977, 979 (7th Cir. 1997)

Heckert v. 5470 Investments Ltd. [2008] B.C.S.C. 1298

Hunter v Southam [1984] 2 S.C.R. 145 at 155

Jones v. Tsige, 2012 ONCA 32, 2012-01-18

Katz v. United States, 389 U.S. 347, 361 (1967)

Kirch v. Embarq Mgmt. Co., 702 F.3d 1245, 1248-49 (10th Cir. 2012)

Kyllo v. United States, 533 U.S. 27 (2001)

Olmstead v. United States, 277 U.S. 438 (1928).

R. v. Dyment [1988] 2 S.C.R. 417

R v Gomboc [2009] ABCA

R v. Plant [1993] 3 S.C.R. 281

Southco, Inc. v. Kanebridge Corp., 390 F.3d 276, 285-87 (3d Cir. 2004)

Whalen v Roe 429 US589

Interview with doctor Ann Cavoukian, former Information and Privacy Commissioner for Ontario (1997-2014) She is currently Executive Director of the Privacy and Big Data Institute at Ryerson University.

Dt. Cavoukian, what is your conception of modern privacy in a big data world?

I don't believe in a 0 sum model, so that you can have privacy or disclosure. I want to have both: privacy and data utility. The way you can do that, especially in data analytics is de-identifying the personal identifiable data. We are moving not only in the direct identifiers but also the indirect identifiers in a manner that enables you to significantly minimize the risk of re-identification.
When you do that, I think that it would also incentivize disclosure of data. It is the removal of the risk associated with disclosure of PII, because they won't be personally identifiable anymore. Therefore removing the privacy harms associated with any potential disclosure or analytics relating to the data.

So you are not a supporter of Paul Ohm theories about the risks of a massive wave of re-identification.

The reason why I reject this view about the grade of risk is that it's nonsense. The problem is that legal scholars do not have the necessarily skills to do that, to assess whether you can truly de identify data and minimize risk. Let's say 0,1%, dramatically minimizing risk. And this, in my opinion, the most expectable view, because there is no 0 level risk. Those who fear re-identification would like to nullify the risk, but it is simply not possible. They want to reduce risk to 0 risk but there's no 0 risk anywhere in the world. Why would we expect that in this area? However, if you can dramatically minimize the risk of re-identification by using very strong de identification protocols, I think that absolutely you have to do that. All the articles that came out critical to de identification, fearing the risks of re identification. They are dealing with problems related to weak de- identification at the beginning.
If the de-identification is poor, weak, of course we will have a greater risk of re-identification. The fact that a weak encryption could be decrypted more easily is an obvious thing. And the reason I have a problem with those articles is that this line of thinking undermines the trust in de identification. Recently, four new models of strong de identification came out. They can't just remove the direct identifier; they can do much more than that.
The risk in this case is the same of being hit by lightning when you go out in a rainstorm.

What about the European approach, do you think that the holistic approach, made up of classic privacy concepts and the implementation of privacy by design mechanism is the right way to tackle the big data issue?

Definitely. I love it. I'm thinking of the GDPR, article 23, article 30 that are dealing with data protection through privacy by design, privacy by default. The reason why I love it is that it combines traditional means of regulatory compliance: the information practices etc..., with this new way of proactively protecting privacy by design embedded into the design features of your IT, business and operational practices; in such a manner that you get the best of both worlds. When I developed privacy by design in the late 90's, the reason why I developed it was to unify path at the international framework for privacy. Since 2010, PbD principles have been translated in 37 languages. The reason why these principles are having such a success is that they are complementary to regulatory compliance around the world. My colleagues realized in 2010's that it was no longer possible just to react to regulatory compliance because the majority of the harms were escaping our attention and detection. As privacy regulators we were watching only the tip of the iceberg of huge amount of privacy harms that remained unknown and unchallenged, unregulated. That's why Pbd can be described as a "medical" model of prevention.

May we hold that if the big data phenomenon represents a quantitative shift that comes out to be a qualitative one, privacy by design represents the counterpart of a qualitative shift in the way the protection of privacy is conceived?

Big data and privacy. I believe that it means not only that we have to work to prevent privacy harms, but also that we have to abandon the 0-sum models in terms of how we look at privacy vs data analytics. We have to get rid of the "or" and embrace the "and". You can have both as I said before. In this sense privacy fosters innovation.

And do you think that this conception of privacy will shift the field of studies from legal studies to the technology?

What is sure is that the legal regulatory model is not enough anymore. Every single time I met someone who was interested in privacy by design I repeated it: you can do it, you can embed privacy within the design of the data architecture. You have to take a holistic approach to privacy. You cannot separate worlds saying: technology is over there, lawyers are over there.

The legal issues remain important but we cannot only rely on them.

What about the implementation? Should PbD be conceived as a competitive advantage?

Definitely. I always say to companies: "don't do it because you think it's the right thing to do, treat privacy as a business issue". Doing it you gain a competitive advantage toward your competitors. A sustainable one.

When I think at the threats posed by big data to individual privacy, the first, which comes to my mind, is the possibility of generating sensitive information from the aggregation of non-sensitive ones, as the target case teaches. Do you agree with me?

Yes, however I have to admit that the trend of collecting a growing volume of information won't decrease in the next years. Companies and technology (AI for example) will need to gather information from the world; that's why embedding privacy within the technology is so important.

We have to appeal to businesses at a level of their own self-interest and say them that a public distrust based on a poor privacy shield will give them the opposite of a competitive advantage. And then with governments, we must remain vigilant: we cannot allow forms of mass surveillance. We have to ask for a higher degree of accountability that will result in more transparency.

But we have to know that there are different approaches to private and public sector. In the private sector companies have to tell their customers what they are doing with their personal information.

Just because companies collect their personal information doesn't mean they own it. The collector has to take care of the data and along with that comes a strong duty of care.

And do you agree with those theories, which consider the Canadian approach as a halfway between EU and US?

Canadian system is really close to Europe. We fulfill the white list requirements; we know that now everything change with the GDPR so we have to remain tuned.

What do you think of the big data phenomenon? Just a trend or something more?

I'm pretty sure it is not just a trend. Moreover, the interesting part is its combination, for example with the internet of things.

A last question. Do you see any flaw or weakness in the Pbd system?

You know no perfect system exist. The 7 foundational principles of Pbd are just a base to build structures. But if you start working from that base privacy will be embedded. We

have a lot of work to do. Implement the FIPPs principles, cyber security, end-to-end security.

The Student Paper Series of the Trento Lawtech Research Group is published since Fall 2010

http://www.lawtech.jus.unitn.it/index.php/student-paper-series?start=1

Freely downloadable papers already published:

STUDENT PAPER N. 25

Leggere (nel)la giurisprudenza: 53 sentenze inedite in tema di responsabilità civile nelle analisi di 53 annotatori in formazione = Reading (in) the caselaw: 53 unpublished judgemnts dealing with civil liability law analyzed with annotations and comments by 53 students during their civil law course

REMO ANDREOLLI, DALILA MACCIONI, ALBERTO MANTOVANI, CHIARA MARCHETTO, MARIASOLE MASCHIO, GIULIA MASSIMO, ALICE MATTEOTTI, MICHELE MAZZETTI, PIERA MIGNEMI, CHIARA MILANESE, GIACOMO MINGARDO, ANNA LAURA MOGETTA, AMEDEO MONTI, SARA MORANDI, BENEDETTA MUNARI, EDOARDO NADALINI, SERENA NANNI, VANIA ODORIZZI, ANTONIA PALOMBELLA, EMANUELE PASTORINO, JULIA PAU, TOMMASO PEDRAZZANI, PATRIZIA PEDRETTI, VERA PERRICONE, BEATRICE PEVARELLO, LARA PIASERE, MARTA PILOTTO, MARCO POLI, ANNA POLITO, CARLO ALBERTO PULEJO, SILVIA RICCAMBONI, ROBERTA RICCHIUTI, LORENZO RICCO, ELEONORA RIGHI, FRANCESCA RIGO, CHIARA ROMANO, ANTONIO ROSSI, ELEONORA ROTOLA, ALESSANDRO RUFFINI, DENISE SACCO, GIULIA SAKEZI, CHIARA SALATI, MATTEO SANTOMAURO, SILVIA SARTORI, ANGELA SETTE, BIANCA STELZER, GIORGIA TRENTINI, SILVIA TROVATO, GIULIA URBANIS, MARIA CRISTINA URBANO, NICOL VECCARO, VERONICA VILLOTTI, GIULIA VISENTINI, LETIZIA ZAVATTI, ELENA ZUCCHI (2016) Leggere (nel)la giurisprudenza: 53 sentenze inedite in tema di responsabilità civile nelle analisi di 53 annotatori in formazione = Reading (in) the caselaw: 53 unpublished judgemnts dealing with civil liability law analyzed with annotations and comments by 53

students during their civil law course. The Trento Law and Technology Research Group. Student Paper Series; 25. Trento: Università degli Studi di Trento. ISBN 978-88-8443-626-9

STUDENT PAPER N. 24

La digitalizzazione del prodotto difettoso: stampa 3D e responsabilità civile= The Digital Defective Product: 3D Product and Civil Liability

CAERAN, MIRCO (2016) La digitalizzazione del prodotto difettoso: stampa 3D e responsabilità civile= The Digital Defective Product: 3D Product and Civil Liability. The Trento Law and Technology Research Group. Student Paper Series; 24. Trento: Università degli Studi di Trento. ISBN 978-88-8443-663-4

STUDENT PAPER N. 23

La gestione della proprietà intellettuale nelle università australiane = Intellectual Property Management in Australian Universities

CHIARUTTINI, MARIA OTTAVIA (2015) La gestione della proprietà intellettuale nelle università australiane = Intellectual Property Management in Australian Universities. The Trento Law and Technology Research Group. Student Paper Series; 23. Trento: Università degli Studi di Trento. ISBN 978-88-8443-626-9

STUDENT PAPER N. 22

Trasferimento tecnologico e realtà locale: vecchie problematiche e nuove prospettive per una collaborazione tra università, industria e territorio = Technology Transfer and Regional Context: Old Problems and New Perspectives for a Sustainable Co-operation among Universitiy, Entrepreneurship and Local Economy

CALGARO, GIOVANNI (2013) Trasferimento tecnologico e realtà locale: vecchie problematiche e nuove prospettive per una collaborazione tra università, industria e territorio. The Trento Law and Technology Research Group. Student Paper Series; 22. Trento: Università degli Studi di Trento. ISBN 978-88-8443-525-5

STUDENT PAPER N. 21

La responsabilità dell'Internet Service Provider per violazione del diritto d'autore: un'analisi comparata = Internet Service Provider liability and copyright infringement: a comparative analysis.

Imperadori, Rossella (2014) *La responsabilità dell'Internet Service Provider per violazione del diritto d'autore: un'analisi comparata.* Trento Law and Technology Research Group. Student Paper; 21 . Trento : Università degli Studi di Trento. ISBN 978-88-8443-572-9

STUDENT PAPER N. 20

Open innovation e patent: un'analisi comparata = Open innovation and patent: a comparative analysis

Ponti, Stefania (2014) *Open innovation e patent: un'analisi comparata.* The Trento Law and Technology Research Group. Student Paper Series; 20 . Trento : Università degli Studi di Trento. ISBN 978-88-8443-573-6

STUDENT PAPER N. 19

La responsabilità civile nell'attività sciistica

CAPPA, MARISA (2014) La responsabilità civile nell'attività sciistica = Ski accidents and civil liability. Trento Law and Technology Research Group. Student Paper Series, 19. Trento: Università degli Studi di Trento.

STUDENT PAPER N. 18

Biodiversità agricola e tutela degli agricoltori dall'Hold-Up brevettuale: il caso degli OGM

TEBANO, GIANLUIGI (2014) Biodiversità agricola e tutela degli agricoltori dall'Hold-Up brevettuale: il caso degli OGM = Agricultural Biodiversity and the Protection of Farmers from patent Hold-Up: the case of GMOs.Trento Law and Technology Research Group. Student Paper Series; 18. Trento : Università degli Studi di Trento.

STUDENT PAPER N. 17

Produrre e nutrirsi "bio" : analisi comparata del diritto degli alimenti biologici

MAFFEI, STEPHANIE (2013) Produrre e nutrirsi "bio" : analisi comparata del diritto degli alimenti biologici = Producing and Eating "Bio": A Comparative Analisys of the Law of Organic Food. Trento Law and Technology Research Group. Student Paper Series; 17. Trento : Università degli Studi di Trento.

STUDENT PAPER N. 16

La tutela delle indicazioni geografiche nel settore vitivinicolo: un'analisi comparata = The Protection of Geographical Indications in the Wine Sector: A Comparative Analisys

SIMONI, CHIARA (2013) La tutela delle indicazioni geografiche nel settore vitivinicolo: un'analisi comparata. The Trento Law and Technology Research Group. Student Papers Series; 16. Trento: Università degli Studi di Trento. Facoltà di Giurisprudenza.
This paper is published in the Trento Law and Technology Research Group - Student Paper Series Electronic copy available at: http://eprints.biblio.unitn.it/archive/00004292/
142

STUDENT PAPER N. 15

Regole di sicurezza e responsabilità civile nelle attività di mountain biking e downhill montano

SALVADORI, IVAN (2013) Regole di sicurezza e responsabilità civile nelle attività di mountain biking e downhill montano. Trento Law and Technology Research Group. Student Paper; 15. Trento: Università degli Studi di Trento.

STUDENT PAPER N. 14

Plagio, proprietà intellettuale e musica: un'analisi interdisciplinare

VIZZIELLO, VIVIANA (2013) Plagio, proprietà intellettuale e musica: un'analisi interdisciplinare. Trento Law and Technology Research Group. Student Paper; 14. Trento: Università degli Studi di Trento.

STUDENT PAPER N.13

The Intellectual Property and Open Source Approaches to Biological Material

CARVALHO, ALEXANDRA (2013) The Intellectual Property and Open Source Approaches to Biological Material. Trento Law and Technology Research Group. Student Paper Series; 13. Trento: Università degli Studi di Trento.

STUDENT PAPER N.12

Per un'archeologia del diritto alimentare: 54 anni di repertori giurisprudenziali sulla sicurezza e qualità del cibo (1876-1930)

TRESTINI, SILVIA (2012) Per un'archeologia del diritto alimentare: 54 anni di repertori giurisprudenziali sulla sicurezza e qualità del cibo (1876-1930) = For an Archeology of

Food Law: 54 Years of Case Law Collections Concerning the Safety and Quality of Food (1876-1930). The Trento Law and Technology Research Group. Student Papers Series, 12. This paper is published in the Trento Law and Technology Research Group - Student Paper Series Electronic copy available at: http://eprints.biblio.unitn.it/archive/00004292/ 143

STUDENT PAPER N.11

Dalle Alpi ai Pirenei: analisi comparata della responsabilità civile per attività turistico-ricreative legate alla montagna nel diritto italiano e spagnolo

PICCIN, CHIARA (2012) Dalle Alpi ai Pirenei: analisi comparata della responsabilità civile per attività turistico-ricreative legate alla montagna nel diritto italiano e spagnolo = From the Alps to the Pyrenees: Comparative Analysis of Civil Liability for Mountain Sport Activities in Italian and Spanish Law. The Trento Law and Technology Research Group. Student Papers Series, 11

STUDENT PAPER N.10

Copynorms: Norme Sociali e Diritto d'Autore

PERRI, THOMAS (2012) Copynorms: Norme Sociali e Diritto d'Autore = Copynorms: Social Norms and Copyright. Trento Law and Technology Research Group. Students Paper Series, 10

STUDENT PAPER N. 9

L'export vitivinicolo negli Stati Uniti: regole di settore e prassi contrattuali con particolare riferimento al caso del Prosecco

ALESSANDRA ZUCCATO (2012), L'export vitivinicolo negli Stati Uniti: regole di settore e prassi contrattuali con particolare riferimento al caso del Prosecco = Exporting Wines to the United States: Rules and Contractual Practices with Specific Reference to the Case of ProseccoTrento: Università degli Studi di Trento (Trento Law and Technology Research Group. Students Paper Series 9)

STUDENT PAPER N.8

Equo compenso e diritto d'autore: un'analisi comparata = Fair Compensation and Author's Rights: a Comparative Analysis.

RUGGERO, BROGI (2011) Equo compenso e diritto d'autore: un'analisi comparata = Fair Compensation and Author's Rights: a Comparative Analysis. Trento: Università degli Studi di Trento (TrentoLawand Technology Research Group. Student Papers Series, 8)
This paper is published in the Trento Law and Technology Research Group - Student Paper Series Electronic copy available at: http://eprints.biblio.unitn.it/archive/00004292/ 144

STUDENT PAPER N.7

Evoluzione tecnologica e mutamento del concetto di plagio nella musica

TREVISA, ANDREA (2012) Evoluzione tecnologica e mutamento del concetto di plagio nella musica = Technological evolution and change of the notion of plagiarism in music Trento: Università degli Studi di Trento (Trento Law and Technology Research Group. Students Paper Series 7)

STUDENT PAPER N.6

Il trasferimento tecnologico università-imprese: profili giuridici ed economici

SIRAGNA, SARA (2011) Il trasferimento tecnologico università-imprese: profili giuridici ed economici = University-Enterprises Technological Transfer: Legal and Economic issues Trento: Università degli Studi di Trento (Trento Law and Technology Research Group. Students Paper Series 6)

STUDENT PAPER N.5

Conciliare la responsabilità medica: il modello "generalista" italiano a confronto col modello "specializzato" francese

GUERRINI, SUSANNA (2011) Conciliare la responsabilità medica: il modello "generalista" italiano a confronto col modello "specializzato" francese = Mediation & Medical Liability: The Italian "General Approach" Compared to the Specialized Model Applied in FranceTrento: Università degli Studi di Trento (Trento Law and Technology Research Group. Students Paper Series 5)

STUDENT PAPER N.4

"Gun Control" e Responsabilità Civile: una comparazione fra Stati Uniti e Italia

PODETTI, MASSIMILIANO (2011) "Gun Control" e Responsabilità Civile: una comparazione fra Stati Uniti e Italia = Gun Control and Tort Liability: A Comparison between the U.S. and Italy Trento: Università degli Studi di Trento. (Trento Law and Technology Research Group. Students Paper Series 4)

145

STUDENT PAPER N.3

Smart Foods e Integratori Alimentari: Profili di Regolamentazione e Responsabilità in una comparazione tra Europa e Stati Uniti

TOGNI, ENRICO (2011) Smart Foods e Integratori Alimentari: Profili di Regolamentazione e Responsabilità in una comparazione tra Europa e Stati Uniti = Smart Foods and Dietary Supplements: Regulatory and Civil Liability Issues in a Comparison between Europe and United States Trento: Università degli Studi di Trento - (Trento Law and Technology Research Group. Students Paper Series; 3)

STUDENT PAPER N.2

Il ruolo della responsabilità civile nella famiglia: una comparazione tra Italia e Francia

SARTOR, MARTA (2010) Il ruolo della responsabilità civile nella famiglia: una comparazione tra Italia e Francia = The Role of Tort Law within the Family: A Comparison between Italy and France Trento: Università degli Studi di Trento - (Trento Law and Technology Research Group. Students
Paper Series; 2)

STUDENT PAPER N.1

Tecnologie belliche e danno al proprio combattente: il ruolo della responsabilità civile in una comparazione fra il caso statunitense dell'Agent Orange e il caso italiano dell'uranio impoverito

RIZZETTO, FEDERICO (2010) Tecnologie belliche e danno al proprio combattente: il ruolo della responsabilità civile in una comparazione fra il caso statunitense dell'Agent Orange e il caso italiano dell'uranio impoverito = War Technologies and Home Soldiers Injuries: The Role of Tort Law in a Comparison

between the American “Agent Orange” and the Italian “Depleted Uranium” Litigations Trento: Università degli Studi di Trento - (Trento Law and Technology Research Group. Students Paper Series; 1)