IMPACT: Journal of Digital Information Technology (IMPACT: JDIT)
Vol. 1, Issue 1, Jun 2016, 1-6

BIG DATA: INTELLECTUAL PROPERTY AND LEGAL ISSUES

YASSIR ELRAYAH

College of Technology / Eastern Michigan University / Michigan, US

ABSTRACT

This paper discussed the legal aspect and challenges of Big Data as an emerging technology. The paper explored how Big Data with its physical and information architecture fits in terms of privacy and intellectual property rights. The study explored many privacy challenges including how Big Data blur the boundaries between government, business and individuals. In relation to Big Data patent and copyright protection, the study concluded that there are lack of legal framework or litigation to resolve legal issues. Furthermore, a legal Big Data model is needed in order manage risk and become an essential factor for the success of the Big Data projects.

KEYWORDS: Big Data, Copyright, Patent, Trademark, Legal

INTRODUCTION

"Big Data" emerged as a new technology with enormous social and economic values. During the last decade technology advancement in terms of computing power, storage capacity, analytics and the high volume and the availability of information gave rise to the "Big Data", which made several businesses, governments, and the public to be more interested in this technology.

WHAT IS BIG DATA?

Data evolved from small scale to large scale after unlimited channels became available. Information grows each year and it will grow in volume by 40% per year as noted by McKinsey Global Institute (Manyika, 2011). The source of Big Data varies and originates from everywhere. Climate information gathered using sensors, posts to the social media websites, financial and transactional records are just a few examples (Munir, 2015). There are several definitions of Big Data, which may differ based on the context. The McKinsey Global Institute (MGI) defined Big Dataas "datasets whose size is beyond the ability of typical database tools to capture, store, manage and analyze." (Manyika, 2011). Tech American Foundation defined Big Data as “high velocity, complex and variable data that require advanced techniques and technologies to enable the capture, storage, distribution, management, and analysis of the information “(Navetta, 2013). Tech American Foundation outlined Big Data by three factors: volume, velocity, and variety. Others include veracityas a fourth factor. The word volume refers to the data or the data density that must be analyzed or managed. Velocity refers to how fast the data is being generated or produced or the speed of the processed data. This will enable real-time data analysis and quick decision-making. Variety refers to the different data processed or analyzed such as a structured, unstructured, text-based, internal and external type of data. Last, veracity refers to the source of the data and whether the collected data are accurate and the source is trusted. (Elster, 2013).

In the business context, Big Data means that an organization must access inconceivable amount of structured and unstructured data. In addition, the organization must realize that they will obtain valuable insights by capturing, structuring

and analyzing these massive volumes of data and therefore improve how an organization does business. The organization must also leverage data analysis tools and data scientist employees to capture, cure, store, search, analyze and share these valuable data. Additionally, the organization need to address potential limitations, security and legal risks associated with the processing of Big Data (Navetta, 2013).

BIG DATA IN LEGAL TERMS

Before discussing the legal terms of Big Data, nature and the purpose of data and information should be defined. Data as information such as messages or common observation which is not confined to boundary or limits are not subjected to legal rules or ownership. However, legal terms and aspects are considered when data is represented as digital information on a disk or database software. Therefore, the rights and duties are in relation to data storage and not to the data itself. These rights and duties are conspicuous through intellectual property rights (IP). There are two levels of the legal framework of Big Data. The first level is the physical infrastructure, which includes hardware such as servers, devices and networks, and software such as operating systems and database applications. The second level is the information architecture such as database schema, data model and database architecture (Kemp, 2014).

BIG DATA AND PRIVACY CHALLENGES

Big Data provides enormous opportunity for economic growth in many fields including national security, medical research, engineering, and technology to name a few. However, Big Data privacy concerns have grown both publicly and privately. The widely noticed privacy concerns come from blurring the boundaries between government, business and individual, which may lead to social problems such as racial profiling, discrimination or freedom restrictions (Polonetsky, 2013). There are tremendous challenges to privacy in the domain of Big Data due to many reasons including the availability of data through different avenues such as mobile devices that have the capability of tracking locations. In addition, the advancement in the infrastructure such as the availability of high-speed transfer networks and robust operating systems. The large storage capacity and the high-efficiency data processors besides cloud computing and the computational frameworks such as Apache Hadoop (Munir, 2015) are all reasons, which influences the privacy of Big Data. Another challenge to Big Data privacy is the ability to extract and interpret data from data such as data coming from the internet logs, surveillance cameras, mobile phones and credit and debit cards (Munir, 2015). Often, Big Data enables identification of data subjects using non-personal data. This will weaken anonymization, which will cast doubt on the essential distinction between personal data and non-personal data (Ohm, 2010, P.57). There are also concerns about automated decision making through automated processes in Big Data about individual's life such as benefits eligibility, credit ratings, or job prospects (Omer, 2013). The ability to combine and mining data will provide more and deep information about individual lives.

INTELLECTUAL PROPERTY OVERVIEW

Intellectual property refers to the creations of people. The law provides intellectual property rights to those people who create their unique and original work in order to protect their creations from any infringement such as work distribution or copying the work without authorization. The list of protected subject matter includes literary, artistic and scientific works, inventions, industrial designs and trademarks. Intellectual property relates to the knowledge or information and not to the tangible media or objects it was incorporated in. To protect the intellectual property of Big Data,

there are different mechanisms and challenges associated with them (Dogra, 2012). Mechanisms such as copyright, patent, and trademark will be discussed in relation to Big Data and their challenges.

BIG DATA ANDCOPYRIGHT PROTECTION ISSUES

Copyright refer to the exclusive legal given right to the creator of the work by law to use or distribute the work for a limited time. Therefore, the author has exclusive rights to his creation and such rights include preventing production or rights to make copies. In the United States, electronic datasets or as known as compilations are protected under U.S copyright Act (1976). Compilation, as defined by the Act, is " collection and assembling of preexisting materials or data that are selected, coordinated, or arranged in such a way that the resulting work constitutes an original work of authorship" (17 U.S.C § 101). In relation to data**,** copyright protects the expression of information and not the original information itself. This expression includes software and certain databases. To prove copyright infringement, the claim must proof that the work is original and not copied from elsewhere. Moreover, the claimant is the owner of the work or can sue on that copyright and the work is within the copyright period, and there is a copyright infringement. Database copyright differs from software copyright; the standard of originality is higher. On 1991, Supreme Court ruled in *Feist Publications v. Rural Telephone Service Co*. (1991) that white page telephone directories were not copyrightable. The main reason was that the compilations were not sufficiently original for protection. The implicit meaning of the ruling is the following. For a compilation such as a white page directory or a factual database to be protected by copyright, it must have a minimal creative selection or arrangement process. Therefore, part of the author's intellectual creation is to arrange and select the database contents. In addition, data protection includes dataset aggregation even if the data is part of the public domain and therefore, not confidential (Kemp, 2014).

Birnhack noted that for a very large dataset, the database is usually unstructured. Therefore, there is no data arrangement or a particular form. To protect Big Data database under copyright Act, the selection, and the arrangement must be creative, which is not always the case for unstructured Big Data database. The database controller should consider alternative protection such as using technological protection measures. Another alternative is to apply trade secret law to keep data confidential (Birnhack, 2013). One of the challenges to protect Big Data is to assess and review risks such as data source, the term, and use of both structured and unstructured data and any contract or license terms. There are many reasons for little litigation in the Big Data domain. In the financial sector processing market data happen at great speed, which makes the evidential burden informal dispute resolution to show what happened, time-consuming and costly(Kemp, 2014). In addition, organizations that rely on data mining do not know in advance, what they may discover, therefore, they cannot provide adequate notice or authorization or cannot knowingly consent to the use of their data mining for Big Data analytics (Rubinstein, 2013)

PATENT PROTECTION

In the domain of Big Data new technologies is used every day to extract useful information from that data. In addition, these technologies sometimes combine data obtained from multiple and different sources. Open source project Hadoop is one of the best examples of Big Data analysis tool that works with the cloud services (Cumbley, 2013). Unlike to structured related databases, Big Data is usually unstructured or semi-structured. From an infrastructure perspective, very few tools can process Big Data. Therefore, researchers are working to develop processing platform for Big Data and this would drive innovation. However, there is still lack of patent litigation that addresses industrial intellectual

property(Sheng, 2012).

Big Data innovations require a disciplined approach in both development and management. One of the issues, which is related to patents in Big Data, is that algorithms are abstract ideas and do not fulfill the eligibility requirements for patent protection. Patent protection is for processes that are novel, non-obvious and useful. Another limitation is the definiteness requirements, so patent claim, must be outlined indefinite terms. This rule made Federal Circuit to invalidate patents claiming processes that rely on subjective judgments. In *Datamize LLV v. Plumtree Software Inc.* (2005), the court patent claims failed the indefiniteness requirement, because it relied on the subjective opinion of a person performing the claimed invention(Mattioli, 2014). However, some lawyers proposed intellectual property protection framework for Big Data. For example, Kappos, proposed a framework, which comprises the four-step process. The first step is to facilitate meetings with Big Data stakeholder and focus group. Then to create invention ideas workshops. Then to address intellectual property law issues and finally, to develop best practices and templates for Big Data, which is related to invention, gathering and evaluation (Kappos, 2014)

TRADEMARK PROTECTION

Trademark protection grants the owner the exclusive right to use them to identify goods or services. The main trademarks intellectual property rights are usually applied to data product and not to the actual data. Data is normally considered separate from the service or product, which it is marketed. The owner who limits that trademark cannot prevent the third party from the ' honest use' of its marks if it was used in a descriptive way. For example in *Golden Nugget, Inc. v American Stock Exchange, Inc* (1987), the American Stock Exchange (ASE), issued, listed and traded options without Golden Nuggets consent. Golden Nugget's stock claimed that ASE had misappropriated its property and infringed Golden Nugget trademark. The US circuit court found Golden Nugget gave up any proprietary rights in the shares having sold its shares to the public (Kemp, 2014).

CONCLUSIONS

Big Data defined by its four factors, volume, variety, velocity, and veracity, opened new avenues for managing and analyzing unstructured data. Big Data become common every day. Companies use Big Data to collect, analyze large dataset to advance their business. Big Data analytics became widely used in many organizations. Despite the extensive use, however, there are legal challenges in privacy and intellectual property protection. For instance, a legal framework that protects personal and business data in Big Data is absent. In addition, a legal model to clearly define and explain the rights and duties for Big Data will not only manage risk but will also become an essential factor for successful Big Data projects. A data protection with its different forms including copyright, patent or trademark guideline for Big Data should be adopted by organizations. However, data protection should strike the right balance between threats and opportunities.

REFERENCES

1. Birnhack, M., (2013*).* S-M-L-XL Data: Big Data as a New Informational Privacy Paradigm, in Big Data and Privacy: Making Ends Meet.
2. Cumbley, R., Church, P., (2013). Is “Big Data” creepy?Computer law & security review 29, 601-609. London, UK

3. Copyright Act of 1976, 17 U.S.C. § 102 *et seq.* (1976).
4. Daysart, J., (2013). The Dawn of Big Data, ABA Journal, 99, 42-45, 47.
5. Dogra, S. (2012). Single Signon For Virtual Machines (VMs) in an Administrative Domain. Technical report.
6. DATAMIZE, LLC v. Plumtree Software, Inc., 417 F. 3d 1342 (2005)
7. Elster, J., (2013). Big Data for law firms, Legal Management,
8. Feist Publications, Inc., v. Rural Telephone Service Co**.**, 499 U.S. 340 (1991)
9. Granstr and, O, Holgersson, M. (2015). Intellectual Property The Wiley-Blackwell Encyclopedia of Consumption and Consumer Studies
10. Golden Nugget, Inc. v American Stock Exchange, Inc828 F.2d 586 (1987)
11. Géczy, P., (2014). Big Data Characteristics. The Macro theme Review 3(6)
12. Hashem, I., Yaqoob, I., NorBadrulAnuar, Mokhtar, S., AbdullahGani, SameeUllahKhan, (2014), The rise of "Big Data" on cloud computing: Review and open research issues. Information Systems.
13. Information commissioner's office, (1998). Big Data and data protection.
14. International Journal of Social, Education, Economics, and Management Engineering.
15. Manyika, J., Chui, M., Brown, B., Bughin, J., R. Dobbs, C. Roxburgh, and A. H. Byers (2011), Big data: The next frontier for innovation, competition, and productivity.
16. Kuner, C., Cate, F., Millard, C. and. Svantesson, D. (2012). The Challenge of 'Big Data' for data protection. International Data Privacy Law, Vol. 2, No. 2
17. Kappos, D. (2014). Intellectual Property Implications and Opportunities in Big Data. New York Law Journal, volume 251—NO. 34.
18. Kemp, R. (2014). Legal aspects of managing Big Data. Computer law & security review 30, 482 -491. London. UK.
19. Little, L. (2013). Big Data: Legal Rights and Obligations
20. Munir A., Yasin, S., Muhammad-Sukki, F. (2015). Big Data: Big Challenges to Privacy and Data Protection, World Academy of Science, Engineering and Technology
21. Murphy M., Barton J. (2014). From a sea of data to actionable insights: Big Data and what it means for lawyers. Intellectual property & Technology Law Journal, volume 26
22. Mattioli, M., (2014), Disclosing Big Data.
23. Ohm P., (2010), Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization,(P.57)
24. Omer, T. and Polonetsky, J. (2013). Big Data for All: Privacy and User Control in theAge of Analytics. Northwestern Journal of Technology and Intellectual Property

25. Quest, P., (2015), Big Data-Big Challenges, Big Opportunities, Pro Quest information report

26. Rubinstein, I., (2013). Big Data: The End of Privacy or a New Beginning. International Data Privacy Law

27. Stough, R., McBride, D. (2014). Big Data and U.S. Public Policy, Review of Policy Research, Volume 31, Number 4 (2014) 10.1111/ropr.12083.

28. Sheng, F., Wang, K. (2015). Field Study of Patent Strategies from Patent Map on Big Data:An Empirical Case of Big Data Application Platformin Taiwan, Cloud System Software Institute, The Institute for Information Industry.

29. Truyens, M., Eecke, P.,(2014). Legal aspects of text mining, computer law & security review 30

30. Tene1, O. and Polonetsky, J. (2013). Big Data for All: Privacy and User Control in the Age of Analytics, Northwestern Journal of Technology and Intellectual Property, Vol 11

31. Winkels, R., (1998), The Open laws project: Big Open Legal Data, University of Amsterdam, Leibniz Center for Law.

32. Wilks, J., Christie, A. (2013). Rights in Data handbook, Protecting and exploiting IP in data and databases international, January 2013.