Hey Siri, Where's My Data? *

The Use of Social Media Exhaust Data

Kevin Angotti†
College of Science
Bellevue University
Bellevue NE USA
kjangotti@my365@bellevue.edu

Ragunath Gunasekaran
College of Science
Bellevue University
Bellevue NE USA

Jennifer Ruiz
College of Science
Bellevue University
Bellevue NE USA
jfruiz@my365.bellevue.edu

ABSTRACT

For most people born in the last three generations, Social Media plays a significant role in their everyday lives. Photos, causes, status updates, and tweets are shared nearly every second. Per 2014 ACI information, every minute, Facebook users share 2.5 million pieces of content; Twitter users tweet nearly three million times [9]. But what happens to all the information that becomes part of one's social media profile?

Our research focused on the current state of social media exhaust data, its secondary uses, and processing challenges of the data. The ethical dilemmas surrounding exhaust data were also examined with particular consideration for user privacy.

CCS CONCEPTS

• General Literature • Applied Computing • Human-Computer Interaction

KEYWORDS

Social Media, Exhaust Data, Data Management, Ethical Usage

ACM Reference format:

Kevin Angotti, Ragunath Gunasekaran, and Jennifer Ruiz. 2020. Hey Siri, Where's My Data? The Use of Social Media Exhaust Data. *ACM, Bellevue, NE, USA, 5 pages.*

1 Introduction

In the last decade, social media has found its way into every aspect of most people's lives. We now live in the age of the "digital human" [23]. Its reaches can be found in shopping habits, social norms, and even in relationship choices. Companies across the world have found usefulness not only for the captured data from users but their exhaust data as well. Exhaust Data is defined as "data that are a by-product of a process whose primary purpose is something other than data capture" [11]. For every image shared, tweeted, retweeted, or liked, a range of exhaust data is generated: who shared, who viewed, what device was used, and the time of day. [24].

Companies around the globe use exhaust data from these sources as a new form of currency, one that provides insights into the lives of those who generate the data [4]. This data serves as a valuable commodity for organizations beyond their originators. Often, data is exchanged between sources and used to further business objectives for those sources [5].

As Mark Cote so eloquently points out, "it is an old adage that if something is free, it must be you that is being sold" [22]. Companies such as Facebook, with over 500 million users, harness the data generated to create new engagement initiatives [18]. Exhaust data provide enough data points to allow for predictive modeling of user activity, both on social media platforms and in real-world activities [15].

The bulk of the discourse on Big Data's downsides today addresses threats to personal information privacy. Loss of privacy can result in other harms, such as identity theft and cyberbullying or cyberstalking [2]. In this article, we examine several areas surrounding exhaust data from social media sources and provide observations for future usage.

1.2 Processing Challenges

Data Scientists and IT professionals have long grappled the challenge of processing and storing the enormous amounts of data generated daily through social media [7, 8]. Challenges include:

1. Storage
2. Scrubbing identifying information
3. Access Control

1.3 Privacy

The use of exhaust data from social media applications comes with several privacy concerns. Considerations must be made to control for imputed identity when triangulating data from multiple sources, as is standard practice [3]. Unauthorized access to user data, as seen with data breaches in recent years or through social engineering tactics, jeopardize privacy, and user trust [10]. Location tracking, both through apps and smart devices, presents a significant threat to privacy [19]. Privacy laws dealing with this area have not yet caught up with the rate of development, leaving many considerations unaddressed.

*Hey Siri, Where's My Data? The Use of Socia Media Exhaust Data
†Kevin Angotti, Ragunath Gunasekaran, & Jennifer Ruiz

1.4 Ethical Considerations

The primary ethical concern raised by the use of exhaust data from social media is that of user consent. Much of the data collected is done so autonomously, and its purpose for usage is ambiguous [21]. Strong concerns also lie in the monetization of such data, as well as the negative impact on public good that it presents [20]. The secondary ethical concern lies in the use of location data. Said data is collected even when users are not actively engaged in social media applications [16]. A final consideration is given to behavioral modeling derived from exhaust data. Use of such data to influence public opinion or polarize a minority group negatively impacts public trust [17].

2 Why Is This Data Science?

Data is currently engulfed throughout our everyday lives. Data about data or metadata, as it is known, can be used to help in decision making and automation. Metadata is all around us, and it is present all the time; we can't shut it off. Nearly every device uses it or generates it, or both [1]. What is the cost of the amount of data we produce with the technology of today? Using online search engines, websites, and social media platforms, we create a byproduct from that data called exhaust data. This type of data can be used for many different things; it is for the benefit of the user or a company, which means that it can either be used for good or bad. Different aspects of exhaust data must be examined to make that determination.

A 2013 study conducted at the University of Stanford attempted to replicate the NSA's PRISM data collection of phone metadata. The study concluded that a tremendous amount of information could be collected from only the metadata. [6]. The advancements of data collection in data science allow for personal information and or National data to be retrieved. Though the use of data science is a tremendous step in the right direction, there are ways to exploit this technology. To show an example from the Stanford study, "a study participant who called "a home improvement store, locksmiths, a hydroponics dealer, and a head shop." Perhaps this individual had perfectly innocent reasons for placing all of these calls, and perhaps these calls were entirely unrelated, but that's not the inference that most of us are likely to make" [6]. From this law enforcement would have a valid reason to open an investigation on that person as a risk to Nation Security. The counter argument to this as a rights violation is that the protection of National Security should be placed above the rights of this individual.

How we use social media today allows for mass amounts of exhaust data to be retrieved and used for the benefit of others. Each social media platform enables users to use the platform in different ways, but they all filter the same thing, data. We use these platforms for friendships, keeping in touch, pictures, or simply just putting your thoughts out for all to see. Each of these platforms allow companies to gather data, and this collection occurs as a result of actions, you, the user are doing. Exhaust data is information-seeking behavior, which can be used to infer people's needs, desires, or intentions [3]. You provide access to this data by way of their user agreements. What you as the user do determines how much data can be retrieved. This is the exhaust data collected, and while they are designed to be free services that help facilitate connectivity among users, they have become more like platforms that deploy strategies to get its users to produce more and more data. The data collected will be analyzed for profits. Moreover, the results from such platforms are sold to advertisers and marketers [16].

2.1 Objectives

This research addresses three specific objectives. First, it looks to analyze the current uses of Exhaust Data generated from social media sources. Second, we seek to discuss the current ethical concerns and provide recommendations to manage these concerns. Lastly, we seek to suggest future uses of Exhaust Data from Social Media.

2.2 Deliverables

Our primary objective is to identify current uses of exhaust data generated from social media sites. Research regarding this usage, as well as the ethics surrounding it, have not been appropriately investigated. The authors also see to provide solutions regarding the ethical and management challenges of usage for incremental improvement.

Data Exhaust can yield some valuable insights for businesses.

1. Improve Market Research: Knowing the ideal audience and customer is essential to effective marketing and product development. While some portions of market research must be done manually (such as surveys and focus groups), others just involve getting into the analytics.
2. Help Shore Up Cybersecurity: a growing threat worldwide, cybercrime has the potential to cost businesses millions. Exhaust Data can also be used to assess risk in different Databases to create a cybersecurity plan.
3. Information Marketing and Product Development strategies: Exhaust Data can help focus on what is important. For example, if your leftover data tells you that most people visit your site via a mobile device, it might be worth considering developing an app to serve those customers [13].
4. Data Exhaust from Manufacturers: The most prominent producers of Data Exhaust are manufacturers and retailers. Standard members of this class include manufacturers of appliances, vehicles, and equipment, and big box stores. They act as receivers of vast amounts of data daily. The majority of this data is used to make changes in the next model, from the user experience to engine parts. The remaining data may not be relevant to the task at hand but could be of use to a supplier or

partner company. Data that can help with quality control or product performance may not be of value to the first receiver but could be crucial to a secondary receiver.

5. Data Exhaust is a Revenue Opportunity: In a highly competitive industry such as automotive manufacturing, it is in the interest of manufacturing leads to promote the use of Big Data through the supply chain for product improvements, quality assurance, cost efficiencies, and innovation. This model can have significant economic benefits to the manufacturer and supply chain. Even the supply chain can sell the data to their supply chain or others who may be interested. This creates a whole new value chain. Exhaust Data can be invaluable to others and help defray your Data center costs [12].

6. Racing to Win with Data: Despite considerable growth in data from IoT devices, only a small amount (8.6 zettabytes) is sent to data centers for storage and subsequent analysis. The 'Data Exhaust' is much bigger than what is analyzed for insights. In artificial intelligence (AI), more of the exhaust data can be analyzed for new insights and turned into real-time actions [14]. Self-driving cars are a great example of harvesting large amounts of sensor data to learn safe and efficient driving behaviors based on each new scenario and environment.

7. Responsible (and legal) use of Data Exhaust: having Data Exhaust at your disposal can be advantageous for a business. However, it is imperative to be cautious and to use Data Exhaust responsibly. There are possible legal implications; it is recommended that organizations consult with a professional regarding the appropriate ways to use the data the organization has accumulated.

Unfortunately, it is also possible to alienate consumers using exhaust data. A recent example is an insurance company using GPS tracking driving habits and charging more to individuals who park in high crime areas. This practice can become discriminatory. Businesses need to be aware of the implications for how they use data to avoid adverse outcomes. When used correctly and responsibly, big data can streamline operations and result in a more effective business strategy.

3 Conclusion

One of the primary jobs of a data scientist is to use data to solve complex problems in the real world. The data generated from information posted on Social Media sites provide a wealth of data points to solve problems. These could include everything from crime prevention, consumer behavior, targeted marketing, and many others.

In summary, if IoT is the New internet and Data is the new oil, don't let IoT Data Exhaust go to waste. It just might contain the insights to fuel the competitive advantage that helps to win the future race in your market. But first, it requires building the capabilities to capture and analyze this data efficiently and revamping innovation processes that leverage the power of IoT-generated Data into new products, services, and business models that separate an organization from the pack.

3.1 Observations for the Future

As technology continues to develop, and the reach of social media applications expands, there are several areas that the authors feel will need to be addressed. First, appropriate legislation to safeguard privacy will need to be enacted. Second, the movement toward an "internet of things" disassociates the data from the user and could lead to ethical concerns. Third, the use of social media data to impact the greater good has to be paramount. Finally, user consent needs to be re-examined so that transparent disclosure of data usage is understood.

ACKNOWLEDGMENTS

The authors would like to begin by thanking Dr. Shankar Parajulee for his knowledge and support throughout the project. They would also like to thank our classmates for their feedback and input. They further like to thank Twitter, Instagram, and all the other apps that Facebook owns for the excellent research idea. Finally, they would like to thank the countless data scientists around the world who further our knowledge and drive our curiosity each day.

REFERENCES

[1] Carson, A. L. 1. acarson@iastate. ed., & Ou, C. carol. ou@unlv. ed. (2019). Notes on Operations Metadata Revisited: Updating Metadata Profiles and Practices in a Vendor-Hosted Repository. Library Resources & Technical Services, 63(4), 196–205.

[2] Markus, M. L. (2015). New games, new rules, new scoreboards: the potential consequences of big data. *Journal of Information Technology (Palgrave Macmillan)*, *30*(1), 58–59. https://doi-org.ezproxy.bellevue.edu/10.1057/jit.2014.28

[3] George, G., Haas, M., & Pentland, A. (2014). Big data and management. *Academy of Management Journal*, (2). Retrieved from https://search-ebscohost-com.ezproxy.bellevue.edu/login.aspx?direct=true&db=edsbig&AN=edsbig.A367721667&site=eds-live

[4] Harford, T. (2014), Big data: A big mistake?. Significance, 11: 14-19.Retrieved from https://doi.org/10.1111/j.1740-9713.2014.00778.x

[5] Gupta, M., George, J.F. (2016). Toward the development of a big data analytics capability. Inf Manag 53(8):1049–1064.

[6] Pomerantz, J. (2015). Metadata. Cambridge, Massachusetts: The MIT Press. https://eds.b.ebscohost.com/eds/detail/detail?vid=1&sid=280fbeab-d30f-4330-8c03-3e0122217708%40sessionmgr101&bdata=JnNpdGU9ZWRzLWxpdmU%3d#AN=belle.2028415&db=cat04787a

[7] Blanas, S. (2018). Towards Exascale Scientific Metadata Management, eScholarship. University of California.

[8] SHUN'ICHI, A. CLAIRE, D., FILIP, M. XML Schema Mappings: Data Exchange and Metadata Management.

[9] Journal of the ACM. Apr2014, Vol. 61, Issue 2, p12-48. 48p.

[10] Eiras. J.R. (2012). Data collection and storage [electronic resource] / editor, New York. Nova Science Publishers, Inc.

[11] Kitchin, R. (2014). The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences. https://methods.sagepub.com/Book/the-data-revolution

[12] Guida, T. (2019). Big Data and Machine Learning in Quantitative Investment. Published by John Wiley and Sons, Ltd.

[13] Shalin, H. Social Media Data Extraction and Content Analysis. Kansas State University, USA. Publisher: IGI Global USA, ISBN: 978-1-5225-0648-5

[14] Danneman, N., Heimann, R. R: Mining spatial, text, web, and social media data Kindle Edition.

[15] Benetka, J., Krumm, J., Bennett, P. (2019). Understanding Context for Tasks and Activities. *The Fourth ACM SIGIR Conference on Human Information Interaction and Retrieval (CHIIR).* Published by ACM https://www.microsoft.com/en-us/research/publication/understanding-context-for-tasks-and-activities/

[16] Langlois, G., Redden, J., Elmer, G. (2015). Compromised Data: From Social Media to Big Data. New York: Bloomsbury Publishing, Inc.

[17] Mergel I., Rethemeyer, R.K., Isett, K. (2016). Big data in public affairs. *Public Administration Review* 76(6): 928–937.

[18] Bughin, J., Chui, M., Manyika, J. (2010). Clouds Big Data and Smart Assets: Ten Tech-Enabled Business Trends to Watch.

[19] Thatcher, J. (2014). Living on fumes: Digital footprints, data fumes, and the limitations of spatial big data. International Journal of Communication 8: 1765–1783.

[20] Markus, M.L. (2015). New games, new rules, new scoreboards: the potential consequences of big data. J Inf Technol 30(1):58–59

[21] Nunan, D., & Di Domenico, M. (2013). Market research and the ethics of big data. International Journal of Market Research, 55(4), 505. doi:10.2501/IJMR-2013-015.

[22] Cote, M. (2014). Data motility: The materiality of big social data. Cultural Studies Review, 20(1), 121–149.

[23] T. H. Rashidi, A. Abbasi, M. Maghrebi, S. Hasan, T. S. Waller. (2017). "Exploring the capacity of social media data for modelling travel behaviour: Opportunities and challenges", Transp. Res. C Emerg. Technol., vol. 75, pp. 197-211.

[24]https://s3.us-east-1.amazonaws.com/blackboard.learn.xythos.prod/5a31d48b683a8/13665053?response-content-disposition=inline%3B%20filename%2A%3DUTF-8%27%27AtlasTechnicalUserGuide.pdf&response-content-type=application%2Fpdf&X-Amz-Algorithm=AWS4-HMAC-SHA256&X-Amz-Date=20191223T002527Z&X-Amz-SignedHeaders=host&X-Amz-Expires=21600&X-Amz-Credential=AKIAIL7WQYDOOHAZJGWQ%2F20191223%2Fus-east-1%2Fs3%2Faws4_request&X-Amz-Signature=b45b58454d9f13bedd9bbcf54a2473d3fd12b53534a14d7353af9545c1ef5d73

[25] Compromised Data: From Social Media to Big Data edited by Greg Elmer, Ganaele Langlois, Joanna Redden, ISBN-13: 978-1501306518, ISBN-10: 1501306510

[26] Twitter and Society (Digital Formations) Edited by Katrin Weller, Axel Bruns, Jean Burgess, Merja Mahrt, Cornelius Puschmann, ISBN-13: 978-1433121692, ISBN-10: 1433121697.

[27] The Influence of Technology on Social Network Analysis and Mining edited by Tansel Özyer, Jon Rokne, Gerhard Wagner, Arno H.P. Reuser.

[28] M. Goetz, J. Leskovec, M. Mcglohon, and C. Faloutsos. (2009). "Modeling Blog Dynamics," AAAI Conference on Weblogs and Social Media.

[29] G. Kossinets, J. Kleinberg, and D. Watts. (2009). "The structure of information pathways in a social communication network," KDD.

[30] P. Melville, W. Gryc, and R. Lawrence, "Sentiment Analysis of Blogs by Combining Lexical Knowledge with Text Classification," KDD, 2009.

[31] Azure Media Services' new AI-powered innovation by Sudheer Sirivara General Manager, Azure Media.